(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94376 A1

(51) International Patent Classification[7]: C07H 21/041, C12Q 1/68

(21) International Application Number: PCT/US01/18815

(22) International Filing Date: 11 June 2001 (11.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/210,884 9 June 2000 (09.06.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ANASTASIO, Alison, E.** [US/US]; 367 Alden Avenue #8B, New Haven, CT 06515 (US). **DUDA, Amy** [US/IT]; Via delle Forze Armate, 260/12, I-20152 20152 Milano (IT). **KLIEM, Stefanie, E.** [DE/US]; 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **SAUSKER, Elizabeth, Ann** [US/US]; 2122 Avalon Haven Drive, North Haven, CT 06473 (US).

(74) Agents: **SHANER, Sandra, L.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/94376 A1

(54) Title: HAPLOTYPES OF THE CFL1 GENE

(57) Abstract: Novel genetic variants of the Cofilin 1 (Non-Muscle) (CFL1) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the CFL1 gene. Compositions and methods for haplotyping and/or genotyping the CFL1 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE CFL1 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/210,884 filed June 9, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human cofilin 1 (non-muscle) (CFL1) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other

groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of immunological disorders is the cofilin 1 (non-muscle) (CFL1) gene or its encoded product. CFL1 is a widely distributed intracellular actin-modulating protein that binds and depolymerizes filamentous F-actin and inhibits the polymerization of monomeric G-actin in a pH-dependent manner. It is involved in the translocation of the actin-cofilin complex from cytoplasm to nucleus. CFL1 is also involved in a pathway through which costimulatory signals for human T-cells are transduced (Nebl et al., *J Biol Chem.* 1996; 271:26276-26280). Costimulation leads to dephosphorylation of CFL1, which subsequently

accumulates in the T-cell nucleus. Inhibition of CFL1 dephosphorylation by okadaic acid reduces its concentrations in the nuclei and is accompanied by apoptosis.

The continuous polymerization and depolymerization of actin filaments is required for various cell responses such as the formation of lamellipodia, filopodia, and chemotaxis (Lian et al., *J Biol Chem.* 2000; 275:2869-2876). Quantitatively, CFL1 is one of the major phosphoproteins in unstimulated neutrophils. Upon stimulation with the chemoattractant fMet-Leu-Phe (fMLP), neutrophils exhibit a rapid and complete dephosphorylation of CFL1 along with a massive translocation of this protein to the actin-rich, ruffling membranes. Thus, because of its involvement with T-cells and neutrophils, CFL1 may play an important role in the rearrangements of the actin cytoskeleton during immunological processes.

The cofilin 1 (non-muscle) gene is located on chromosome 11q13 and contains 4 exons that encode a 166 amino acid protein. A reference sequence for the CFL1 gene is shown in Figure 1(Genaissance Contig No. 2140102; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_005507) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Buetow et al., (Nat Genet. 1999 Mar;21(3):323-5) identified a polymorphism of cytosine or thymine in the CFL1 gene at a position corresponding to nucleotide 27701 in Figure 1.

Because of the potential for variation in the CFL1 gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the CFL1 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of CFL1 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 12 novel polymorphic sites in the CFL1 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 25247 (PS1), 25319 (PS2), 25336 (PS3), 25398 (PS4), 25445 (PS5), 25624 (PS6), 25758 (PS7), 27481 (PS8), 27566 (PS9), 27873 (PS11), 27980 (PS12) and 28164 (PS13). The polymorphisms at these sites are guanine or thymine at PS1, cytosine or thymine at PS2, guanine or adenine at PS3, adenine or guanine at PS4, cytosine or thymine at PS5, guanine or cytosine at PS6, cytosine or thymine at PS7, thymine or cytosine at PS8, thymine or guanine at PS9, cytosine or thymine at PS11, guanine or adenine at PS12 and cytosine or thymine at PS13. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified site at nucleotide position 27701 (PS10) in Figure 1, in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS13 in the CFL1 gene, which are shown below in Tables 4 and 3, respectively. Each of these

CFL1 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the CFL1 gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the CFL1 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13 in both copies of the CFL1 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel CFL1 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel CFL1 polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-PS13. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the CFL1 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the CFL1 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's CFL1 gene is defined by one of the CFL1 haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's CFL1 gene are defined by one of the CFL1 haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. The method for establishing the CFL1 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with CFL1 activity, e.g., immunological disorders .

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate CFL1 as a candidate target for treating a specific condition or disease predicted to be associated with CFL1 activity. Determining for a particular population the frequency of one or more of the individual CFL1 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue CFL1 as a target for treating the specific disease of interest. In particular, if variable CFL1 activity is associated with the disease, then one or more CFL1 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed CFL1 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable CFL1 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any CFL1 haplotype or haplotype pair, apply the information derived from

detecting CFL1 haplotypes in an individual to decide whether modulating CFL1 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting CFL1 to treat a specific condition or disease predicted to be associated with CFL1 activity. For example, detecting which of the CFL1 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent CFL1 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular CFL1 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the CFL1 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with CFL1 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the CFL1 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute CFL1 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a CFL1 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any CFL1 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a CFL1 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the CFL1 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the CFL1 genotype or haplotype in a reference population. A higher frequency of the CFL1 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the CFL1 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the CFL1 haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for immunological disorders .

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the CFL1 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 (SEQ ID NO:1) and the polymorphic variant comprises at least one polymorphism selected from the group

consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, thymine at PS5, cytosine at PS6, thymine at PS7, cytosine at PS8, guanine at PS9, thymine at PS11, adenine at PS12 and thymine at PS13. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of thymine at PS10.

A particularly preferred polymorphic variant is an isogene of the CFL1 gene. A CFL1 isogene of the invention comprises guanine or thymine at PS1, cytosine or thymine at PS2, guanine or adenine at PS3, adenine or guanine at PS4, cytosine or thymine at PS5, guanine or cytosine at PS6, cytosine or thymine at PS7, thymine or cytosine at PS8, thymine or guanine at PS9, cytosine or thymine at PS10, cytosine or thymine at PS11, guanine or adenine at PS12 and cytosine or thymine at PS13. The invention also provides a collection of CFL1 isogenes, referred to herein as a CFL1 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a CFL1 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises guanine at a position corresponding to nucleotide 63. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of thymine at a position corresponding to nucleotide 198. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CFL1 isogene defined by haplotypes 2, 8, 11, 14 and 15.

Polynucleotides complementary to these CFL1 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the CFL1 gene will be useful in studying the expression and function of CFL1, and in expressing CFL1 protein for use in screening for candidate drugs to treat diseases related to CFL1 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express CFL1 for protein structure analysis and drug binding studies.

The present invention also provides nonhuman transgenic animals comprising one of the CFL1 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the CFL1 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against CFL1 protein, and for testing the efficacy of therapeutic agents and compounds for immunological disorders in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the CFL1 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the CFL1 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing CFL1 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the CFL1 gene (Genaissance Reference No. 2140102; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:64 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:65 is a modified version of SEQ ID NO:64 that shows the context sequence of each polymorphic site, PS1-PS13, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:65 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the CFL1 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the CFL1 protein (contiguous lines; SEQ ID NO:3).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the CFL1 gene. As described in more detail below, the inventors herein discovered 15 isogenes of the CFL1 gene by characterizing the CFL1 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The CFL1 isogenes present in the human reference population are defined by haplotypes for 13 polymorphic sites in the CFL1 gene, 12 of which are believed to be novel. The CFL1 polymorphic sites identified by the inventors are referred to as PS1-PS13 to designate the order in which they are located in the gene (see Table 2 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13. Using the genotypes identified in the Index Repository for PS1-PS13 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the CFL1 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the CFL1 gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether CFL1 is a suitable target for drugs to treat immunological disorders , screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods

of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the CFL1 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel CFL1 polymorphisms and haplotypes identified herein.

The compositions comprise at least one CFL1 genotyping oligonucleotide. In one embodiment, a CFL1 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a CFL1 polynucleotide, i.e., a CFL1 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-CFL1 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the CFL1 gene using the polymorphism information provided herein in conjunction with the known sequence information for the CFL1 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or

"complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting CFL1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTGCTGTKGGATAAA (SEQ ID NO:4) and its complement,
CTCAGAGYGGTTCCT (SEQ ID NO:5) and its complement,
GAAATTGRACCAATG (SEQ ID NO:6) and its complement,
TCCCTACRTGCACTG (SEQ ID NO:7) and its complement,
TACCCAGYGAGCGAG (SEQ ID NO:8) and its complement,
GCTCTCGSTGCCCTC (SEQ ID NO:9) and its complement,
AGCGCCCYGGGCGGG (SEQ ID NO:10) and its complement,
CTTTTCTYGACGTAT (SEQ ID NO:11) and its complement,
AGGTGCGKAAGTCTT (SEQ ID NO:12) and its complement,
GGAACTTYTGTGGCT (SEQ ID NO:13) and its complement,
TGGCTTTRCTGTTGC (SEQ ID NO:14) and its complement, and
AATGGCAYGCAGAGG (SEQ ID NO:15) and its complement.

A preferred ASO primer for detecting CFL1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCCGTTTTGCTGTKG (SEQ ID NO:16); CGGCTCTTTATCCMA (SEQ ID NO:17);
GCCGGGCTCAGAGYG (SEQ ID NO:18); TTTCCCAGGAACCRC (SEQ ID NO:19);
TCCTGGGAAATTGRA (SEQ ID NO:20); CGAGCCCATTGGTYC (SEQ ID NO:21);
TACATTTCCCTACRT (SEQ ID NO:22); GAGCTGCAGTGCAYG (SEQ ID NO:23);
CTCCGTTACCCAGYG (SEQ ID NO:24); CGCCGCCTCGCTCRC (SEQ ID NO:25);
TCTGCGGCTCTCGST (SEQ ID NO:26); AAAGGAGAGGGCASC (SEQ ID NO:27);
TCCCCCAGCGCCCYG (SEQ ID NO:28); CCTCATCCCGCCCRG (SEQ ID NO:29);
AGAATCCTTTTCTYG (SEQ ID NO:30); CCACGTATACGTCRA (SEQ ID NO:31);
ACATGAAGGTGCGKA (SEQ ID NO:32); GCGTTGAAGACTTMC (SEQ ID NO:33);
ATGATGGGAACTTYT (SEQ ID NO:34); ACCAGGAGCCACARA (SEQ ID NO:35);
CACCTGTGGCTTTRC (SEQ ID NO:36); CTCCCAGCAACAGYA (SEQ ID NO:37);
AGGGTGAATGGCAYG (SEQ ID NO:38); and AGACCCCCTCTGCRT (SEQ ID NO:39).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting CFL1 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GTTTTGCTGT (SEQ ID NO:40); CTCTTTATCC (SEQ ID NO:41);
GGGCTCAGAG (SEQ ID NO:42); CCCAGGAACC (SEQ ID NO:43);
TGGGAAATTG (SEQ ID NO:44); GCCCATTGGT (SEQ ID NO:45);
ATTTCCCTAC (SEQ ID NO:46); CTGCAGTGCA (SEQ ID NO:47);
CGTTACCCAG (SEQ ID NO:48); CGCCTCGCTC (SEQ ID NO:49);
GCGGCTCTCG (SEQ ID NO:50); GGAGAGGGCA (SEQ ID NO:51);
CCCAGCGCCC (SEQ ID NO:52); CATCCCGCCC (SEQ ID NO:53);
ATCCTTTTCT (SEQ ID NO:54); CGTATACGTC (SEQ ID NO:55);

TGAAGGTGCG (SEQ ID NO:56); TTGAAGACTT (SEQ ID NO:57);
ATGGGAACTT (SEQ ID NO:58); AGGAGCCACA (SEQ ID NO:59);
CTGTGGCTTT (SEQ ID NO:60); CCAGCAACAG (SEQ ID NO:61);
GTGAATGGCA (SEQ ID NO:62); and CCCCCTCTGC (SEQ ID NO:63).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

CFL1 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized CFL1 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the CFL1 gene in an individual. As used herein, the terms "CFL1 genotype" and "CFL1 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the CFL1 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the CFL1 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13 in the two copies to assign a CFL1 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair at PS10 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS13.

Typically, the nucleic acid sample is isolated from a biological sample taken from the

individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the CFL1 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If a CFL1 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the CFL1 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13 in that copy to assign a CFL1 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the CFL1 gene or fragment such as one of the methods described above for preparing CFL1 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two CFL1 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional CFL1 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the CFL1 gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at PS10. In a particularly preferred embodiment, the nucleotide at each of PS1-PS13 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the CFL1 haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's CFL1 gene, the phased sequence of nucleotides present at each of PS1-PS13. The present invention also contemplates that typically only a subset of PS1-PS13 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* **97**:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* **22**:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* **4**:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a CFL1 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13 in each copy of the CFL1 gene that is present in the individual. In a particularly preferred embodiment, the

haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS13 in each copy of the CFL1 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the CFL1 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting

temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the CFL1 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in

Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's CFL1 haplotype pair is predicted from its CFL1 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a CFL1 genotype for the individual at two or more CFL1 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing CFL1 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the CFL1 haplotype pairs shown in Table 3.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective

pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a CFL1 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*). A preferred process for predicting CFL1 haplotype pairs from CFL1 genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application filed April 18, 2001.

The invention also provides a method for determining the frequency of a CFL1 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel CFL1 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for CFL1 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a CFL1 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype

pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular CFL1 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that CFL1 genotype, haplotype or haplotype pair. Preferably, the CFL1 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting CFL1 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a CFL1 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not

actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the CFL1 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and CFL1 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their CFL1 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the CFL1 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between CFL1 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the CFL1 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with

one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of CFL1 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the CFL1 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the CFL1 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying CFL1 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the CFL1 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant CFL1 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, and may also comprise an additional polymorphism of thymine at PS10. Similarly, the nucleotide sequence of a variant fragment of the CFL1 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the CFL1 gene, which is defined by haplotype 6, (or other reported CFL1 sequences) or to portions of the reference sequence (or other reported CFL1 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, thymine at PS5, cytosine at PS6, thymine at PS7, cytosine at PS8, guanine at PS9, thymine at PS11, adenine at PS12 and thymine at PS13. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the CFL1 gene which is defined by any one of haplotypes 1- 5 and 7 - 15 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the CFL1 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

CFL1 isogenes may be isolated using any method that allows separation of the two "copies" of the CFL1 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides CFL1 genome anthologies, which are collections of CFL1 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A CFL1 genome anthology may comprise individual CFL1 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the CFL1 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred CFL1 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded CFL1 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant CFL1 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell

using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the CFL1 gene will produce CFL1 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a CFL1 cDNA comprising a nucleotide sequence which is a polymorphic variant of the CFL1 reference coding sequence shown in Figure 2. Thus, the invention also provides CFL1 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having guanine at a position corresponding to nucleotide 63, and may also comprise an additional polymorphism of thymine at a position corresponding to nucleotide 198. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a CFL1 isogene defined by haplotypes 2, 8, 11, 14 and 15. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphism described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized CFL1 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a CFL1 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the CFL1 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the CFL1 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the CFL1 genomic variants described

herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular CFL1 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the CFL1 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular CFL1 isogene. Expression of a CFL1 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of CFL1 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of CFL1 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

Effect(s) of the polymorphisms identified herein on expression of CFL1 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the CFL1 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into CFL1 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired CFL1 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the CFL1 isogene is introduced into a cell in such a way that it recombines with the endogenous CFL1 gene

present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired CFL1 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the CFL1 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the CFL1 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant CFL1 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the CFL1 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human CFL1 isogene and producing human CFL1 protein can be used as biological models for studying diseases related to abnormal CFL1 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel CFL1 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel CFL1 isogenes; an antisense oligonucleotide directed against one of the novel CFL1 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel CFL1 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel CFL1 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered

saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the CFL1 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The CFL1 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the CFL1 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the CFL1 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	25037-25060	complement of 25649-25628	613 nt
Fragment 2	25378-25401	complement of 26056-26037	679 nt
Fragment 3	27289-27311	complement of 27785-27762	497 nt
Fragment 4	27452-27474	complement of 27976-27954	525 nt
Fragment 5	27856-27879	complement of 28258-28238	403 nt
Fragment 6	28115-28135	complement of 28660-28640	546 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec. }

70°C - 45 sec.
72°C - 45 sec. 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	25122-25142	complement of 25622-25603
Fragment 2	25422-25441	complement of 25989-25970
Fragment 3	27363-27385	complement of 27757-27737
Fragment 4	27511-27530	complement of 27950-27931
Fragment 5	27890-27909	complement of 28238-28219
Fragment 6	28182-28201	complement of 28615-28596

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the CFL1 gene are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the CFL1 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position[b]	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	2140145	25247	G	T		
PS2	2140147	25319	C	T		
PS3	2140149	25336	G	A		
PS4	2140153	25398	A	G		
PS5	2140157	25445	C	T		
PS6	2140161	25624	G	C		
PS7	10366962	25758	C	T		
PS8	2140167	27481	T	C		
PS9	2140173	27566	T	G	63	R21R
PS10[R]	2140175	27701	C	T	198	D66D
PS11	2140177	27873	C	T		
PS12	2140179	27980	G	A		
PS13	2140181	28164	C	T		

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
[b]Position within Figure 1
[R]Reported in the literature

EXAMPLE 2

This example illustrates analysis of the CFL1 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3(Part1). Genotypes and Haplotype Pairs Observed for CFL1 Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP Pair
1	G	C	G	A	C	G	C	T	T	T	8 8
2	G	C	G	A	C	G	C	T	T	C	6 6
3	G	C	G	A	C	G	C	C	T	C	4 4
4	G	C	G	A	C	G	T	T	T	C	10 10
5	G	C	G	A	C	C	C	T	T	C	3 3
6	G	C/T	G	A	C	G	C	T	T	T	8 14
7	G/T	C	G	A	C	G	C	T	T	T	8 15
8	G	C	G	A	C	G	T/C	T/C	T	C	10 4
9	G	C	G	A	C	G/C	C	T	T/G	T/C	8 2
10	G	C	G/A	A	C	G/C	C	T	T	C	6 1
11	G	C	G	A	C	G	C	T/C	T	T/C	8 4
12	G	C	G	A	C	G	C/T	T	T	C	6 10
13	G	C	G	A	C	G/C	C	T	T	C	6 3
14	G	C	G	A	C	G	C	T	T	C	6 5
15	G	C	G	A/G	C	G	C	T	T	C	6 13
16	G	C	G	A	C	G/C	C	T	T	T/C	8 3
17	G	C	G	A	C/T	G	C	T	T	C	6 12
18	G	C	G	A/G	C	G	C	T	T	T/C	8 13
19	G	C	G	A	C	G	C	T	T	T/C	8 6
20	G	C	G	A	C	G	C	T/C	T	C	6 4
21	G	C	G	A	C	G	T	T	T	C/T	10 11
22	G	C	G	A	C	G	C	T	T	T/C	9 7

Table 3(Part2). Genotypes and Haplotype Pairs Observed for CFL1 Gene

Genotype Number	Polymorphic Sites PS11	PS12	PS13	HAP Pair
1	T	G	C	8 8
2	C	G	C	6 6
3	C	G	C	4 4
4	C	G	C	10 10
5	C	G	C	3 3
6	T	G	C	8 14
7	T	G	C	8 15
8	C	G	C	10 4
9	T/C	G	C	8 2
10	C	G	C	6 1
11	T/C	G	C	8 4
12	C	G	C	6 10
13	C	G	C	6 3
14	C	G/A	C	6 5
15	C	G	C	6 13
16	T/C	G	C	8 3
17	C	G	C	6 12
18	T/C	G	C	8 13
19	T/C	G	C	8 6
20	C	G	C	6 4
21	C/T	G	C	10 11
22	T/C	G	T	9 7

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for

assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In our analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 15 human CFL1 haplotypes shown in Table 4 below.

An CFL1 isogene defined by a full-haplotype shown in Table 4 below comprises the regions of the SEQ ID NOS indicated in Table 4, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 4.

Table 4. Haplotypes Observed in the CFL1 Gene

Haplotype Number[a]															PS Number[b]	PS Position[c]	SEQ ID NO:	Region Examined[d]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15				
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647/30	64/65	3437-4456
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719/150	64/65	3437-4456
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736/270	64/65	3437-4456
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798/390	64/65	3437-4456
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845/510	64/65	3437-4456
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024/630	64/65	3437-4456
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158/750	64/65	3437-4456
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881/870	64/65	5689-7060
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966/990	64/65	5689-7060
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101/1110	64/65	5689-7060
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273/1230	64/65	5689-7060
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380/1350	64/65	5689-7060
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564/1470	64/65	5689-7060

[a]Alleles for CFL1 haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;
[d]1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS13, to facilitate electronic searching of the haplotypes;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:64 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:65 is a modified version of SEQ ID NO:64 that shows the context sequence of each of PS1-PS13 in a uniform format to facilitate electronic searching of the CFL1 haplotypes. For each polymorphic site, SEQ ID NO:65 contains a block of 60

bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 5 below shows the percent of chromosomes characterized by a given CFL1 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given CFL1 haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and NA = Native American.

Table 5. Frequency of Observed CFL1 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	10405275	0.61	0.0	2.5	0.0	0.0	0.0
2	10405276	0.61	0.0	2.5	0.0	0.0	0.0
3	10405271	2.44	0.0	10.0	0.0	0.0	0.0
4	10405269	8.54	16.67	10.0	5.0	2.78	0.0
5	10405277	0.61	0.0	2.5	0.0	0.0	0.0
6	10405267	20.73	7.14	45.0	22.5	5.56	33.33
7	10405274	0.61	0.0	2.5	0.0	0.0	0.0
8	10405266	43.9	69.05	12.5	30.0	61.11	66.67
9	10405273	0.61	0.0	2.5	0.0	0.0	0.0
10	10405268	10.98	2.38	5.0	25.0	13.89	0.0
11	10405270	6.71	2.38	2.5	10.0	13.89	0.0
12	10405278	0.61	0.0	2.5	0.0	0.0	0.0
13	10405272	1.83	0.0	0.0	7.5	0.0	0.0
14	10405279	0.61	0.0	0.0	0.0	2.78	0.0
15	10405280	0.61	2.38	0.0	0.0	0.0	0.0

Table 6. Frequency of Observed CFL1 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
8	8	30.49	47.62	0.0	20.0	50.0	66.67
6	6	6.1	0.0	15.0	5.0	0.0	33.33
4	4	1.22	4.76	0.0	0.0	0.0	0.0
10	10	2.44	0.0	0.0	10.0	0.0	0.0
3	3	1.22	0.0	5.0	0.0	0.0	0.0
8	14	1.22	0.0	0.0	0.0	5.56	0.0
8	15	1.22	4.76	0.0	0.0	0.0	0.0
10	4	1.22	0.0	0.0	5.0	0.0	0.0
8	2	1.22	0.0	5.0	0.0	0.0	0.0
6	1	1.22	0.0	5.0	0.0	0.0	0.0
8	4	7.32	23.81	0.0	0.0	5.56	0.0
6	10	2.44	0.0	5.0	5.0	0.0	0.0
6	3	1.22	0.0	5.0	0.0	0.0	0.0
6	5	1.22	0.0	5.0	0.0	0.0	0.0
6	13	2.44	0.0	0.0	10.0	0.0	0.0
8	3	1.22	0.0	5.0	0.0	0.0	0.0
6	12	1.22	0.0	5.0	0.0	0.0	0.0
8	13	1.22	0.0	0.0	5.0	0.0	0.0
8	6	13.41	14.29	15.0	15.0	11.11	0.0
6	4	6.1	0.0	20.0	5.0	0.0	0.0
10	11	13.41	4.76	5.0	20.0	27.78	0.0
9	7	1.22	0.0	5.0	0.0	0.0	0.0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the CFL1 gene are likely to be similar to the relative frequencies of these CFL1 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved

and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the cofilin 1 (non-muscle) (CFL1) gene of an individual, which comprises determining which of the CFL1 haplotypes shown in the table immediately below defines one copy of the individual's CFL1 gene, wherein each of the CFL1 haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]															PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS within SEQ ID NO:64.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS13 on the one copy of the individual's CFL1 gene.

3. A method for haplotyping the cofilin 1 (non-muscle) (CFL1) gene of an individual, which comprises determining which of the CFL1 haplotype pairs shown in the table immediately below defines both copies of the individual's CFL1 gene, wherein each of the CFL1 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11		
G/G	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	3736
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C	G/C	G/G	6	4024
C/C	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	7	4158
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/C	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/G	T/T	T/T	9	5966
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	10	6101
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	11	6273
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	6564

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
12	13	14	15	16	17	18	19	20	21	22		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	3736
A/A	A/A	A/A	A/G	A/A	A/A	A/G	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/C	G/G	G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	6	4024
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C	7	4158
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	9	5966
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	10	6101
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	11	6273
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	13	6564

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS13 on both copies of the individual's CFL1 gene.

5. A method for genotyping the cofilin 1 (non-muscle) (CFL1) gene of an individual, comprising determining for the two copies of the CFL1 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:64.

6. The method of claim 5, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the CFL1

gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the CFL1 gene present in the individual the identity of the nucleotide pair at each of PS1-PS13.

8. A method for haplotyping the cofilin 1 (non-muscle) (CFL1) gene of an individual which comprises determining, for one copy of the CFL1 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:64.

9. The method of claim 8, further comprising determining the identity of the nucleotide at PS10, which has the location and alternative alleles shown in SEQ ID NO:64.

10. The method of claim 8, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the CFL1 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

11. A method for predicting a haplotype pair for the cofilin 1 (non-muscle) (CFL1) gene of an individual comprising:

(a) identifying a CFL1 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:64;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and

(d) assigning a haplotype pair to the individual that is consistent with the data

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11		
G/G	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	3736
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C	G/C	G/G	6	4024
C/C	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	7	4158
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/C	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/G	T/T	T/T	9	5966
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	10	6101
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	11	6273
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	6564

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
12	13	14	15	16	17	18	19	20	21	22		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	3736
A/A	A/A	A/A	A/G	A/A	A/A	A/G	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/C	G/G	G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	6	4024
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C	7	4158
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	9	5966
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	10	6101
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	11	6273
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	13	6564

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64.

12. The method of claim 11, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS13, which have the location and alternative alleles shown in SEQ ID NO:64.

13. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the cofilin 1 (non-muscle) (CFL1) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the

haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-15 shown in the table presented immediately below, wherein each of the haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]															PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the CFL1 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11		
G/G	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	3736
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C	G/C	G/G	6	4024
C/C	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	7	4158
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/C	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/G	T/T	T/T	9	5966
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	10	6101
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	11	6273
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	6564

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
12	13	14	15	16	17	18	19	20	21	22		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	3736
A/A	A/A	A/A	A/G	A/A	A/A	A/G	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/C	G/G	G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	6	4024
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C	7	4158
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	9	5966
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	10	6101
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	11	6273
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	13	6564

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

14. The method of claim 13, wherein the trait is a clinical response to a drug targeting CFL1.

15. An isolated genotyping oligonucleotide for detecting a polymorphism in the cofilin 1 (non-muscle) (CFL1) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:64.

16. The isolated genotyping oligonucleotide of claim 15, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the CFL1 gene at a region containing the polymorphic site.

17. The allele-specific oligonucleotide of claim 16, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-15, the complements of SEQ ID NOS:4-15, and SEQ ID NOS:16-39.

18. The isolated genotyping oligonucleotide of claim 15, which is a primer-extension oligonucleotide.

19. The primer-extension oligonucleotide of claim 18,which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:40-63.

20. A kit for genotyping the cofilin 1 (non-muscle) (CFL1) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12 and PS13, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:64.

21. The kit of claim 20, which further comprises oligonucleotides designed to genotype PS10,

having the location and alternative alleles shown in SEQ ID NO:64.

22. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a cofilin 1 (non-muscle) (CFL1) isogene, wherein the CFL1 isogene is selected from the group consisting of isogenes 1- 5 and 7 - 15 shown in the table immediately below and wherein each of the isogenes comprises the regions of the SEQ ID NOS shown in the table immediately below and wherein each of the isogenes 1- 5 and 7 - 15 is further defined by the corresponding set of polymorphisms whose locations and identities are set forth in the table immediately below

Isogene Number[a]															PS Number[b]	PS Position[c]	SEQ ID NO:	Region Examined[d]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15				
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647	64	3437-4456
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719	64	3437-4456
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736	64	3437-4456
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798	64	3437-4456
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845	64	3437-4456
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024	64	3437-4456
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158	64	3437-4456
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881	64	5689-7060
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966	64	5689-7060
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101	64	5689-7060
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273	64	5689-7060
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380	64	5689-7060
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564	64	5689-7060

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64;

(b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, thymine at PS2, adenine at PS3, guanine at PS4, thymine at PS5, cytosine at PS6, thymine at PS7, cytosine at PS8, guanine at PS9, thymine at PS11, adenine at PS12 and thymine at PS13, wherein the selected polymorphism has the location set forth in the table immediately above; and

(c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.

23. The isolated polynucleotide of claim 22, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

24. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 22, wherein the organism expresses a CFL1 protein encoded by the first nucleotide sequence.

25. The recombinant nonhuman organism of claim 24, which is a nonhuman transgenic animal.

26. The isolated polynucleotide of claim 22 which consists of the second nucleotide sequence.

27. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a coding sequence for a cofilin 1 (non-muscle) (CFL1) isogene wherein the coding sequence is defined by a haplotype selected from the group consisting of 2, 8, 11, 14 and 15 shown in the table immediately below and wherein the coding sequence comprises SEQ ID NO:2 except at each of the polymorphic sites which have the locations and polymorphisms set forth in the table immediately below:

Coding Sequence Subhaplotype Number[a]		PS Number[b]	PS Position[c]
2c	8c, 11c, 14c, 15c		
G	T	9	63
C	T	10	198

[a]Alleles for coding sequence haplotypes are presented 5′ to 3′ in each column; the numerical portion of the coding sequence haplotype number represents the number of the parent full CFL1 haplotype;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:2;

and

(b) a fragment of the coding sequence, wherein the fragment comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 63, wherein said positions in the coding sequence and the fragment refer to SEQ ID NO:2.

28. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 27, wherein the organism expresses a cofilin 1 (non-muscle) (CFL1) protein encoded by the polymorphic variant sequence.

29. The recombinant nonhuman organism of claim 28, which is a nonhuman transgenic animal.

30. A computer system for storing and analyzing polymorphism data for the cofilin 1 (non-muscle) gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

Haplotype Number[a]															PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64;

and the haplotype pairs set forth in the table immediately below:

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11		
G/G	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	3736
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C	G/C	G/G	6	4024
C/C	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	7	4158
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/C	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/G	T/T	T/T	9	5966
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	10	6101
T/T	C/C	C/C	C/C	C/C	T/T	T/T	C/C	T/C	C/C	T/C	11	6273
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	6564

Haplotype Pair Number[a]											PS Number[b]	PS Position[c]
12	13	14	15	16	17	18	19	20	21	22		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3647
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3719
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	3736
A/A	A/A	A/A	A/G	A/A	A/A	A/G	A/A	A/A	A/A	A/A	4	3798
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	3845
G/G	G/C	G/G	G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	6	4024
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C	7	4158
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	8	5881
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	9	5966
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	10	6101
C/C	C/C	C/C	C/C	T/C	C/C	T/C	T/C	C/C	C/T	T/C	11	6273
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	6380
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	13	6564

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[c]Location of PS in SEQ ID NO:64.

31. A genome anthology for the cofilin 1 (non-muscle) (CFL1) gene which comprises CFL1 isogenes defined by any one of haplotypes 1-15 set forth in the table shown below:

Haplotype Number[a]															PS Number[b]	PS Position[c]
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	1	3647
C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	2	3719
A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	3	3736
A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	4	3798
C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	5	3845
C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	6	4024
C	C	C	C	C	C	C	C	C	T	T	C	C	C	C	7	4158
T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	8	5881
T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	9	5966
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	10	6101
C	C	C	C	C	C	C	T	T	C	T	C	C	T	T	11	6273
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	6380
C	C	C	C	C	C	T	C	T	C	C	C	C	C	C	13	6564

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:64.

POLYMORPHISMS IN THE CFL1 GENE

CAGGTGGCTG GGATAATCTT TGCCTGCCCT GCCTGGGCCA GGCCCCTCGG
GGAACCTCCC ACCCCACTGC CATGGTTCCG CCTCCCGGTG TAGCCCAGTG 21700
CTCACCAGGT GCTCCCGGTA GAGCACCAGC AGTATCACTC GGGCTCCTGA
GTGCCGAGCT GGCCAGTAGC TGCCAGAGCC TCCCGCTTTG GGCTGGGCAG 21800
GAACCTGAAA GGTACAGAGG AGTTCTGAAC ATGAAAGCCA GCCAGTCATC
TTTGAATAAG TCCCCAGCAG GCTTTGTGGC CCTGGGTAAG TCACTCAGTT 21900
GTCTCTCAGT TTCCTTCTCT GTAAAATGGG CCAGGTGATG GGATGATCCG
CTGCAAATCT GGGATAAAAC AGCAGGCAGT TTCGGGGCAG TGGGTGGGGG 22000
TGTCTGCCCA GCCTGGCCTC ACTTCTACCA TGCGCTGCAG GACAGGTGTA
ATTAGACAGG AAAACTCAAT AATGGTGACC CCGTAACTAA CTGAGCTGGG 22100
ATCCTTAAGT AGCTTGATCA TGGAATTCTC TCATTGGCTG TTACCAATAT
TTCTATTCAC AGGTTAAAAT ACTAAGTCGA AAAGAGGATA AGTCATTTGC 22200
TCAAGGCCAG GCATCACTGG TGAGATTTGG ACTCAGGCTT GCCAAATCCC
CCAAGCCAGA AAATCATACT CCCGCACTCC CACTTTGCCT TCCACTCCCC 22300
TTGCCCCTTC CTCACTGGCA TGGAAGAGTC CTGGACTTCC GCAAGTCGCC
CCTGCGGGGC TGGACTGTTC TCTCCAGATG GTGAGTCCGG GGCATGGTCA 22400
CCTGCGGGAG AGAAAAGAGT TAACTCCACA CTCCACGAAT CCGGCAGCTG
CGAGGATCTG GGATGCAACC TACTCAGATG GTTACGTGAG ACCGGTGAGG 22500
GAGGCTCAGG GAGCGCCAGC CACTAGCCCT AGGCCGAACG GCAACTGGGA
CAGGAGTGGG GAGCTTAAAA GCCAAGGGCT CCCAGGAGAG TACGGAGGCT 22600
CCACCTGGGG GAGGTGGGGG ATCACCGACC CCTTTCCGAG GCGCTCACCG
CCCGATGTTC GGTGCCGCTG CAGCCGCTCG TCCAGCATCC GGCAGAGCCC 22700
GTCTCCGAAG TGCTGTAGGA TCTTAGCTTC CTTCCCGCTG CGCAGCGGCA
GTGGGTACCG TCGGAGGGAA CGCAGCGCCT GGCCGGGCAG GGGTAGGGTA 22800
CCTGGTCAGG GCAGGCCTGA CCTGGCCAGC AGCTTTTCCC ATCGGGCCAC
GCCAGGACCC ACCTTCTGAA ATACGAAGCG CGTGCGGCGC CTGCTGCGGG 22900
TCGCCTCGTC CCGCCACTCG GTCAGCCAGC GAACGAAGAG CGGGTTGGGA
CAGGCAGGCA GCGGGCGCTT CCGGCCCAGG CGGACCGGGG CCGCCATGAG 23000
CCCGAGGGCG GGTCCTCCGG CGCTCCACGC CCGCGGGACT GGGACGCCGC
CAGTTCTGGA GTCGGGATTC GAACACCTGG CCCAGGGCGG GGCGGGGGAA 23100
GAGGGAGACC GAGGGCAGGA CCGTTGAGAT CACGGGGCCC CAAACACAGG
GGGCACTAAC GAGAGGAGAG CCTGCGGCCG CCCTCCCACC CCTGCGCTGC 23200
TCTAAGTGGT TGGTCAGGGG CGCTGTCTCT AACGATCTTT TCCCTTGGCT
CCACTCCAGC CAGCCTTTGA GACGCGCCCT TTCCCAGCCC CCGGAGCCTG 23300
GACCGGCTCC TTCCTGTCAC CCCAGATTCG AGAGCGGGGT CTTTGTGTAG
GCGAGAGGAA GGCGGACAGC GCCCCCTGTG ACGGGGAGGA CTGGCTGGGC 23400
GGCGCTTAAC GGACTAACCT GAGCTTGGCG GCTCTAAGGT CTCTTAAGAT
TGTCAGCTTC AGCCTCAAGC TTGAAGGAGG CTGGAGATGG AGACTGGTCT 23500
TGAAAAGAGT CCAGGAGAGG CGGGGGCTAG GAGACAAACG GGTGCAAACG
TATTGGTGGG AGGGACACCC AGAAATGGGT GCCGGGGATG AGGACACAAA 23600
GGTAAGGCAG GAGGAGCTGC AGGACCCTTG GAGGAAATGG TGATCCTCAT
CCCCGGGATA CAGGGCAAAG AGTGTGTGGT GGAGAGGAAG TCGGGGACGT 23700
GCCGAGAAGG CCCAAAATAG AACTCGGAAT GTCGGGGGTC AGAGGCCTAG
AGGAAGGAGG ACTGTAAGTA AGAGGTGCTT GGGCATCCTG GCAGTTGGGG 23800
AGGCCTGAGA TGGGCACGCG GTTGGGTTTG GGAGCGGAGG AGCAGGCTGC
GGGTTAAGAA GCCCAGACAC AGGTAGAGGG GAGAACACTT CGTCATTTTT 23900
GTGGCTAATC GGAAGTGGGA GACCCTCAGG GAGGTTGAAG ATGGAAACCC
TACAAAGGGC TTCAACCTGA GAAGCCTGAA AGACACGTAT AGCTGAGTTG 24000
GACACAGGCG ACCCGCCAGG GTTTGAAGCA GCACCCGGCC TCTAGCTGCT
ATGCCGGAAG TCGCAACACG CAGGTCTCCA GCCTGCGAGG AAGGGAGAGG 24100
TGGGCGCCAG AGCTCCAGTC CCGCCCCTCA GGACTGTGCC TAGCCCACGG

FIGURE 1A

```
CGCAGTAGCC CCAGCCAAGC TAAATTGGAC CAATGGGCTC CGGCTCCCGC      24200
CTCCCGCCCG CCGCGGCCTT ATAAGGAGCC TCAAGCCGGC TGTCGGTCTG
GCCCCGCCCC CTCGAGTCTC TTGTCCTAAT AAGGACATCT AGGGCATCCC      24300
GCGGCTGGGG GCGGGCGAGA GCGCGTCTCC AGGGTAACTG CGAAGCCCCC
CTCTCTCTCC CCGTTGCTCT TGGAAACTCT AAGGCTCAGA AAAGGGGGCG      24400
CTAGCCTCAT ATTCATGTAG GGTCGCTCGC GGGAGGCGGG GTCAAGCTGC
TGTCTCCACC TGGAGCGCCG GGAGTCCGGG CGGGGGCGGT GCTGTGGAAT      24500
GTCCCCGACT ACAACCCCCG GTATGCTCTG CAAGGTACCC CACTCCCAGG
CATCGCTCTC CGTTTACACC CTTAATTCCG TTTATAGGAT TAGGGCCCCC      24600
AGTTCCCTTT CTGGGTTTGC CTCTGACTTG TTTGCGACCC AGTCTTTTCC
TGCAGAGTGT GAGGCTGTGG TCACTGTTCT AGAACAACAA AATCATCCCG      24700
GAGTCCCCCG CTCGTTTCTT TTTTTTTTTA AGGCATATAC TACCACAGAA
CTTGTCTCAG TTTTGCGGAT GTTAGAATAA CCATGTACCT GGCTATCAAC      24800
TCGAGAGCTG GGTTCAAGTC TGGTTTAATA TCTATTTCTT CTCATTGGCA
CGTAAAGGCA TGTAAAGAAG CCTTGAATTG AATGCACAAA CCCTCGTCTT      24900
TACCCTTTAA CGCCTCATGA CTTTGCCAAA CATCTTTAAA CAAGAACAAA
GCTCTCAGAG ATTTCCTTGT ACCTTTCCCC TGTGCCTTTT CCTCCTATTC      25000
TTGTTTTACT GACGTTGGAC ATGGCCTTTT TTCAGGAGAC GAGAGTGGCT
TTACACATGC AGAGTACTAG TGTTTATCCG ACTGACGACG CCTTCATACT      25100
TGTGATTTCC TTTGCTTTAG GTCGGGATTC CAGGCTAATC AATAGTTGCT
TCGATTCAGA AATGCAAAAC CCTAACCTCA CTCAAAAATT TCAGGCCAAG      25200
GGATCCAGAT AATACCAGAT GGGTCTTAAG AAAGCCGTTT TGCTGTGGGA
                                                      T
TAAAGAGCCG CTTAGTCGGG GATCGTTTTC GGGTCATTTT ACTGAGCGCC      25300
GCCTCGCCGG GCTCAGAGCG GTTCCTGGGA AATTGGACCA ATGGGCTCGC
                  T                   A
GCTGCGCGCT GCGGTGCCGC CCAGGACCTG GGCCTACATT TCCCTACATG      25400
                                                    G
CACTGCAGCT CGGAGCCGGC CGGCGGGAAG ACTCCGTTAC CCAGCGAGCG
                                                      T
AGGCGGCGGC GCAGGGCCAG CGGACTCCAT TTCCCGTCGG CTCGCGGTGG      25500
GAGCGCCGGA AGCCCGCCCC ACCCCTCATT GTGCGGCTCC TACTAAACGG
AAGGGGCCGG GAGAGGCCGC GTTCAGTCGG GTCCCGGCAG CGGCTGCAGC      25600
GCTCTCGTCT TCTGCGGCTC TCGGTGCCCT CTCCTTTTCG TTTCCGGAAA
                        C
CATGGTGAGC GGCAGGCCAC GGCGCCTGAG GGAGGCGGCT GCGGGTCGGT      25700
      [exon 1: 25652..
               ..25654]
CGCCTGCGCG CGGGAAGCGA CTGGGGAGCG ACGTCCGCAC CCCCTCCCCC
AGCGCCCCGG GCGGGATGAG GGTCTCGCAC GAAGGGGCGG GCGGTGCCGG      25800
       T
AATGCGCGTG CGCGCCCGTG GGCGCCGGGG AGGGCCGGTC TGGACGTCGC
TCGCGCTCCG CCTGGGCTCC CCTCCCCCAC CCGCTGGTGC GCCCGCGCGG      25900
ACACCGGCCG CGGGGGGAGG GGTTCGGGGC GCGCGCTGGG GCGCCCCTCG
CGGAACGGCC GGCGTCGCGC CTTTCTTCTT AGGGGCGCGC TCTCGACTGG      26000
AGCATGCGGC GTCCAAACCC GGCCCGGGGG CGGCGAGGAA AAAGCGCGCG
AGAGGTCTCG GCGCGCGCCC CACCCAAGGC TGTTCCTGTC TGCGCGTCAC      26100
CCTCCCTCGG CAGCCGGGTC CGTCTCGGGA GCAGGTGGCA GGGATGCCTG
CGCCCGGGAG GGGCGGGTGC TGGCGCGGCG AGGGGTGATG CTGGGAAGAC      26200
GGGGTCCCGG TTGGGGCGTG GGTGCCTGCG ATTAAGGCCT TGCACCGCAA
GGCTTGGAGG GGGCATCGCA GAGACCGCGG CCCGTTCGGG AGCGCATCTA      26300
ACGAGCTGCG TTCTCACCCG TGCATCGGGC GAGGGCTGGA ACGGCGCTGT
CTGTCGGCGC GTGCGCGCAC GCTCAGGCCC GGCCGCCGGT GCCGAAGCCC      26400
```

```
TGGGCCAGCG AGGCCTTCCC GGATGGGCCT GAGTGAGGGT GGAGCCGAGT
TTAGGGAAGT GACCCAGGCG GGCACCGCCC GGAAACCACC CCGCCCCCTG      26500
TATACGGCCC GAGGGCTGAG AAAACCGAAG GTTATGTAGC TTGCCCAGGC
TCCTGCCTGC AGGAGATGGG TCTGATACAC CCGTTGTTTT AGGGCGATTC      26600
CACAGGGTTA GGGACCTGGA GATGCTGCTG CCTCTCGGAT ACGCGTTTCT
GCATTGGTGA GGGGGCCGGG CCCAGCCGAT CTCCTGTCCC CGCCCACTCC      26700
GGATGCGGCC GAGTCACGTG GCCGGCTTCT TCTGCAGTTC CGGGGAGTTT
GGGGGACCAG ATTTACCTTG GATTGCCCCT CCCTCTCCTG GCTCGGGCAC      26800
CCCGAAGCAC GACGCAGAGT AGGAAGAGTT TAAAGACCTC GAGGCCTCTG
GGCACTTGAG TTTGGCATGT TAATTTTTAT CAGCGACTTC TGGGGCCTAG      26900
CACCATTCCC GGAAGAAGGG AGTTGTCGGG CAGGGTCCTT AATGGGGGTT
GCAATTCTTG TCTTGGTTGG GAAAGAGCCT AGCTGGGAAC AGGGGTCGTT      27000
TGTGTAGTAA CTGTATTAAG CAGTTCTGGT GGGTTTGTTG TAACTGTTCG
GTGTCCTTGG GGTCCCATCT AGCATTGTGG GGTCTTGGCT CTTTCTGAAG      27100
AGGAGTCTTG GCTCAGGTGT TTGGGTCCTC CATATCCGAA GTTCTAAATC
ACCAGTTTTA TCCAAAGGCT GAGAGCAAGT ACTGTGAGTA GCAAATGAGT      27200
CAGTAGCCGG TGGATCTTTT AATTCTGGGT TCACATTTTG GTCAAGGTTC
TGCTTTTCAC GAGTTTTGCA TTTCTATCCT GTAAAATGAG GGAAGGGGTT      27300
GTTGAACTAG GTGATCAGAG GCCATGGCTG CCATTCTGAG GTTGTGTCCC
ATCTGGTTCT TTGACAGTGA CTTGGGTGGG AAAACTAGTG ACTTGCCTTG      27400
GGAAGGGCAC TAGTGACCAC TCGGGACGTG GACTGAGATC TTCTGAAAAA
ACAAGGGGCA GTTCCCTAGA ATCCTTTTCT TGACGTATAC GTGGCATGTC      27500
                                  C
TCCTAGGCCT CCGGTGTGGC TGTCTCTGAT GGTGTCATCA AGGTGTTCAA
     [exon 2: 27507..
CGACATGAAG GTGCGTAAGT CTTCAACGCC AGAGGAGGTG AAGAAGCGCA      27600
           G
AGAAGGCGGT GCTCTTCTGC CTGAGTGAGG ACAAGAAGAA CATCATCCTG
GAGGAGGGCA AGGAGATCCT GGTGGGCGAT GTGGGCCAGA CTGTCGACGA      27700
CCCCTACGCC ACCTTTGTCA AGATGCTGCC AGATAAGGAC TGCCGCTATG.
T
CCCTCTATGA TGCAACCTAT GAGACCAAGG AGAGCAAGAA GGAGGATCTG      27800
GTGTTTATCT TCTGGTGAGC TCATTCTGGC ACTTCTTCCT GTCACCTGCC
              ..27814]
CCCTTTCTCA TGATGGGAAC TTCTGTGGCT CCTGGTCAAC CCAGATGTGT      27900
                    T
GGGTTTGGAG GGGGAGGGGG TTGTAACAAA AGACTCCTCA CTGCCAGCCC
AAGGGTTTCT CAAAGTCACC TGTGGCTTTG CTGTTGCTGG GAGTTGCCTG      28000
                                  A
ACGCGTGTTT CTGTCCAGGG CCCCCGAGTC TGCGCCCCTT AAGAGCAAAA
     [exon 3: 28019..
TGATTTATGC CAGCTCCAAG GACGCCATCA AGAAGAAGCT GACAGGTAAG      28100
              ..28095]
GGCCAGCATT GGTGCTGAGT GCCCTTCTGC TCTGGCCTTG GCTGCAGGGC
AGGGTGAATG GCACGCAGAG GGGGTCTGCC CCCTTGTCTT CGCTGCCTGC      28200
           T
TTGTTCCTTC CATCTGCTTT GATTGGCTCC TTCCCAGCCA CAGAAACTAC
CCCCACCCCA GTGCTTCCTG CTCACAGATG CTCCCCTTTC TTCTTTATAG      28300
GGATCAAGCA TGAATTGCAA GCAAACTGCT ACGAGGAGGT CAAGGACCGC
     [exon 4: 28301..
TGCACCCTGG CAGAGAAGCT GGGGGGCAGT GCCGTCATCT CCCTGGAGGG      28400
CAAGCCTTTG TGAGCCCCTT CTGGCCCCCT GCCTGGAGCA TCTGGCAGCC
              ..28413]
```

FIGURE 1C

CCACACCTGC CCTTGGGGGT TGCAGGCTGC CCCCTTCCTG CCAGACCGGA 28500
GGGGCTGGGG GGATCCCAGC AGGGGGAGGG CAATCCCTTC ACCCCAGTTG
CCAAACAGAC CCCCCACCCC CTGGATTTTC CTTCTCCCTC CATCCCTTGA 28600
CGGTTCTGGC CTTCCCAAAC TGCTTTTGAT CTTTTGATTC CTCTTGGGCT
GAAGCAGACC AAGTTCCCCC CAGGCACCCC AGTTGTGGGG GAGCCTGTAT 28700
TTTTTTTAAC AACATCCCCA TTCCCCACCT GGTCCTCCCC CTTCCCATGC
TGCCAACTTC TAACCGCAAT AGTGACTCTG TGCTTGTCTG TTTAGTTCTG 28800
TGTATAAATG GAATGTTGTG GAGATGACCC CTCCCTGTGC CGGCTGGTTC
CTCTCCCTTT TCCCCTGGTC ACGGCTACTC ATGGAAGCAG GACCAGTAAG 28900
GGACCTTCGA TTAAAAAAAA AAAAGACAAT AATAAAAAGG CTCATTAATG
GGATGTGTTT TTCAAGGTTG GGACACAGGA GACTTCAGGA TTGGGGGTGC 29000
ACTGGGATAA TGGTGCTATC TCTCTCCTCT GTACTCCACA CCCAACCAGC
CCCTGCACAC CCTTATCTGG CTCCAATGAC TGGCTGAGAC GCCTTTCTGA 29100
TGGTCACCAT TATTCTGTTG GTGCTGTGAT TGACACTGGA CTGGCTGTTT
AAGGAGTGGG TGCTGGCTAC AGGGGTCTTC CAGCGGAAGT GATTTATGCC 29200
TGAACTGGGT GCTCTGTAGC CTTGCCTGGA TGGACAGTGA GGTCCATCTG
ATATTCTGTT CTGTCCCTTT CTAGGTGTGG GCTGATAGGC AAGTCAGAGG 29300
TAGCTCAAAG GTTCTGTTTG GTCTGCTGTC TTCCCTGGAT GGCTGACTCA
AGTTTATGAG GGTCAGGTTC AGTGGCTGCT GTGAGTCAGG CCAAGTTGAG 29400
CCTTTGACCT CCTCTAGATG TGGCTGGAGT GCATTCAGGG TGGGGCCAGC
CAGGTGGAGC CCGGGACTTC AGACCCCGGT GTCTTGAGCA CTGCCTCCCT 29500
GTGTTTAACC AAAAGGGTGT TGCCGGTACC CTGAGCCTTT TGCCAGCATT
GTCAAGGGCT GTTTCCAGCC TGTACTTTGG GCTTTTTCCT GCCCAGCAGG 29600
CAAAGACATA CAAGAAATCT GGCCCTAGGG TGAGGTGAAC TGGGAGTGGT
AATGTGGCAT CTATATTTCC CTTATCAACT GAACTGAGGA AGAGTCCTCC 29700
TGGTGTTACA GTTGGCTCCA TAACTCCTGG GAACTTGTGA GCGGTTCTGG
GAGGAAGGGC TAGATCCCTT TTCTGGTTAG GAAATAGGTG TGGAGAACTT 29800
GGGATTCATT TTTGTGCTAG GGTTTATGGG GGTAGGGGCA AACCTGGAAG
ACTTCTGGGG TTCCTAGGTC CCAGCTTCTG GGTGTCGCGC CTTCCCACAG 29900
GGAGCCTGCT GTCACTGGAG AGAGGGCTGA GGATGCTGCA GGGGTGCCCT
CCTTGGTGTC TCCAGGATCT GATCATACTG CCTTGGCAGA AAGTGTGCCT 30000
GGTGGACCTT TGAGCTGTGG GAAGGGAAGT AAGGACAACA GGCCTCCTTG
CTACAAAACC AGGGTTTTTA TTTGCGAAGG AGAACAGGCC GGGATCAAGT 30100
GAGCTATGAT TCCTTCAGGG CTATGGGTGA GGGTAGAGGC TGTTCCTCAG
CTTCCTGGGG AGAGGTGTCA GGGTCCCTGG GGCTCTGCTG ACTGTGGCCA 30200
GTTATGATGC CCATGCTGCC CTGTGTCTTG CAGGGCCTGG CACCTGTGGT
GGTGGTGATA GGGGGTTTGG GCTTTCCCTG AGAGGATGTG GGGCAGAGTG 30300
AGGGTGTGGC TAGTGGCACT GCTAGGGAAG GAGACATCTT GCCGGAGAGA
GGGGTCTGGG ATACCCTGGT CACTGGAGAT CCCAGATTAG GGTCTGGCTG 30400
AGCTCTGGCT ACTCTAAGGC TCCCCCTTTA GGGCAACAAG GGTGTTCCGG
AGAATCAGGG TGCTGGCCTG CAGAGGAGAA ACAGGTCCTT GAAGTCCCCT 30500
GCTCCCTGAG TCAGGCTGCC ATTCTTTCTG CACCTTAGAT GCTCCCTCGC
TCCCCACACC CCTGCTCTCT GCAATAACAG CTGCTGTTAA TTGGGCCTTC 30600
CTACCCAGAA GGCACGGGTC CTGAGCCAGG GAAGCATGGC CCCAAGCCTT
AGCTGGGGCT CCTCTCATGC ACACACACGC GCATACATGC GTGCACACCC 30700
TCACATGCAC GCACACCCTG GTGCCTGGGG GAGGGAGGGA GGTGGGAGGT
GGGAGGCTTC ACAAGGGCCC AGTGAGGTGG GAGGGCTCAC CTTGGCGTGT 30800
TTGAGCTCCA GCTCTGCCAG CTCAATGAGA TTCTTTCGAA AAGAGGAGAC
CCGGCGGGAC TTGAAGTCCA TGAGCTCTGG AGAGGAGGGG AAGAGGCATG 30900
GGGATCCGGG CAGGGCAGGG CATGTGGCCC CAGGCTGGGG CTGGGGGCTG
CGGGCTCACC TTGCTTGGCG GAGTCGGAGA GGCGCTCGAA GCGTTGGCAG 31000
CACAGCTGCT GGTGGCTCTC GGCGGGCCGC ACCTCCCGGT TCCTGGTGCG
CGCCTTGTCC AGCGCCTTGT TGGCATTCTC GTAGTCGGCC AGTGCCCGCA 31100

GCCGCCGGTA CAGCAGGTCC TAGGACCATG GGCACAGGGT CACTGCTGCC
CTGTCCTGGC ATCCGGGCCT GAGCGCTGCG CTGGGGCTGC CTCTCACCTT 31200
GGCTGCCTGT GAGTCACGCA TGTAGTACCT CAGCATGTCT GACAGCTTCA
GGTCCTCATC GGAAGCCACC CGGCCCTCCA GCTTCTGCAT CAGAAAGGTG 31300
GGCACTGGGA AGGGAGCTTG TCAGAGGGGC AGCTAAGGGG TGGGAGGCAC
CTCTTCATCT AAATATCCAG CAAAGAAGCC TTAGGTAACT GTTCCAGGTC 31400
AGGTCTGGTG TTGAGGGGTG CAGGCCAGGG AAGGCTTCAC AGAGGAAGTG
GGATCTAGTC TGGATCTCGG GGACTGAGTA AGAGCTGGAC AGAGGGCGCT 31500
GAGGGCTTGG CAGCTGTGCA AGCCCGGCTA TGTGTTCAGG GAAACTTTAG
TTCAGGGAGG GCACTTGGGC AAAGCAGGGG AGGGAGGCAG TGGCTGGAGC 31600
ACAAGAGGCT GAAAGTGGAA GCTCAGGACT TTAGGAGCTT CCAGAAGGGG
CAGGGTCTCT GAAGGGCCTT GAGCAGGGAA GAGATGCGGT CCAATAGACA 31700
TTTTAGGAAG ATGGCAACAC TCCAGTGACT TTCTTAGGGA AGAGTGGGGA
ACATGTGTGC CCGGGCAGTG TGCTGGCAGC TGGGTGCAGC CTGCCAGAGG 31800
TTTGGGCACA CCCTTGTGCC TCCCAACCCT CCGCTGCTGC ACACGTGCAT
ACATGTGCAT GCCCCTGCAC CCCCCAACAC ACACACGAAG CTGGATCCTT 31900
GTCTTCACCC CACCTTTTGT TAAACTAGTC TGGAGTGTAA GCCCAACTTG
CCACTGGGCC CAAGTTATCA CTGAACCCAA GATGATCCTG GAGCTCTTCC 32000
CTCTGCCAGG ACTCATCTTC CCTGCAAAGC TCACCAGGAA AGCCTTTGAG
CACAGAAGGC AGTACTCACC CTCAGCCGTT CAAAGAGCTC TGCCAATTTG 32100
AGGAAGCTCC TAGGAAGCAG TAGGGAGGGA GTTAGGGTAG GGGGCTGAGG
AACAGCTCAA AGCCTCTCTG TCTCAGCCTC TCCTTGGGAA TCTCCCTCTC 32200
CCCTTTCCTT CCCAACACCA TCCCCCAACA GGAGAGCCTA CACCTCCCAC
CTAGAGCTGA CAGTAGACAT CTACAGAGTG ACCCCTCATG GGAAGACATG 32300
GTATTACAGG CACAGGGAGA GGGTAGGGGT GGGGGGAGTC CTCACGTCTT
TAGCTGGTTG ACTTCCTGTG TTCCCAGACT GCTCAGCGCA GCTGAGATAG 32400
GGATATAATC GTC 32413

FIGURE 1E

POLYMORPHISMS IN THE CODING SEQUENCE OF CFL1

```
ATGGCCTCCG GTGTGGCTGT CTCTGATGGT GTCATCAAGG TGTTCAACGA
CATGAAGGTG CGTAAGTCTT CAACGCCAGA GGAGGTGAAG AAGCGCAAGA      100
             G
AGGCGGTGCT CTTCTGCCTG AGTGAGGACA AGAAGAACAT CATCCTGGAG
GAGGGCAAGG AGATCCTGGT GGGCGATGTG GGCCAGACTG TCGACGACCC      200
                                                  T
CTACGCCACC TTTGTCAAGA TGCTGCCAGA TAAGGACTGC CGCTATGCCC
TCTATGATGC AACCTATGAG ACCAAGGAGA GCAAGAAGGA GGATCTGGTG      300
TTTATCTTCT GGGCCCCCGA GTCTGCGCCC CTTAAGAGCA AAATGATTTA
TGCCAGCTCC AAGGACGCCA TCAAGAAGAA GCTGACAGGG ATCAAGCATG      400
AATTGCAAGC AAACTGCTAC GAGGAGGTCA AGGACCGCTG CACCCTGGCA
GAGAAGCTGG GGGGCAGTGC CGTCATCTCC CTGGAGGGCA AGCCTTTGTG      500
A                                                           501
```

AMINO ACID SEQUENCE OF THE CFL1 PROTEIN

MASGVAVSDG VIKVFNDMKV RKSSTPEEVK KRKKAVLFCL SEDKKNIILE
EGKEILVGDV GQTVDDPYAT FVKMLPDKDC RYALYDATYE TKESKKEDLV 100
FIFWAPESAP LKSKMIYASS KDAIKKKLTG IKHELQANCY EEVKDRCTLA
EKLGGSAVIS LEGKPL 166

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Anastasio, Alison E.
Duda, Amy
Kliem, Stefanie E.
Koshy, Beena
Sausker, Elizabeth Ann

<120> Haplotypes of the CFL1 Gene

<130> MWH-0706PCT CFL1

<140> TBA
<141> 2001-06-11

<150> 60/210,884
<151> 2000-06-09

<160> 65

<170> PatentIn Ver. 2.1

<210> 1
<211> 10813
<212> DNA
<213> Homo sapiens

<400> 1
caggtggctg ggataatctt tgcctgccct gcctgggcca ggcccctcgg ggaacctccc 60
accccactgc catggttccg cctcccggtg tagcccagtg ctcaccaggt gctcccggta 120
gagcaccagc agtatcactc gggctcctga gtgccgagct ggccagtagc tgccagagcc 180
tcccgctttg ggctgggcag gaacctgaaa ggtacagagg agttctgaac atgaaagcca 240
gccagtcatc tttgaataag tccccagcag gctttgtggc cctgggtaag tcactcagtt 300
gtctctcagt ttccttctct gtaaaatggg ccaggtgatg ggatgatccg ctgcaaatct 360
gggataaaac agcaggcagt ttcggggcag tgggtggggg tgtctgccca gcctggcctc 420
acttctacca tgcgctgcag gacaggtgta attagacagg aaaactcaat aatggtgacc 480
ccgtaactaa ctgagctggg atccttaagt agcttgatca tggaattctc tcattggctg 540
ttaccaatat ttctattcac aggttaaaat actaagtcga aaagaggata agtcatttgc 600
tcaaggccag gcatcactgg tgagatttgg actcaggctt gccaaatccc ccaagccaga 660
aaatcatact cccgcactcc cactttgcct tccactcccc ttgccccttc ctcactggca 720
tggaagagtc ctggacttcc gcaagtcgcc cctgcggggc tggactgttc tctccagatg 780
gtgagtccgg ggcatggtca cctgcgggag agaaaagagt taactccaca ctccacgaat 840
ccggcagctg cgaggatctg ggatgcaacc tactcagatg gttacgtgag accggtgagg 900
gaggctcagg gagcgccagc cactagccct aggccgaacg gcaactggga caggagtggg 960
gagcttaaaa gccaagggct cccaggagag tacggaggct ccacctgggg gaggtggggg 1020
atcaccgacc cctttccgag gcgctcaccg cccgatgttc ggtgccgctg cagccgctcg 1080
tccagcatcc ggcagagccc gtctccgaag tgctgtagga tcttagcttc cttcccgctg 1140

cgcagcggca gtgggtaccg tcggagggaa cgcagcgcct ggccgggcag gggtagggta 1200
cctggtcagg gcaggcctga cctggccagc agcttttccc atcgggccac gccaggaccc 1260
accttctgaa atacgaagcg cgtgcggcgc ctgctgcggg tcgcctcgtc ccgccactcg 1320
gtcagccagc gaacgaagag cgggttggga caggcaggca gcgggcgctt ccggcccagg 1380
cggaccgggg ccgccatgag cccgagggcg ggtcctccgg cgctccacgc ccgcgggact 1440
gggacgccgc cagttctgga gtcgggattc gaacacctgg cccagggcgg ggcgggggaa 1500
gagggagacc gagggcagga ccgttgagat cacggggccc caaacacagg gggcactaac 1560
gagaggagag cctgcggccg ccctcccacc cctgcgctgc tctaagtggt tggtcagggg 1620
cgctgtctct aacgatcttt tcccttggct ccactccagc cagcctttga gacgcgccct 1680
ttcccagccc ccggagcctg gaccggctcc ttcctgtcac cccagattcg agagcggggt 1740
ctttgtgtag gcgagaggaa ggcggacagc gccccctgtg acggggagga ctggctgggc 1800
ggcgcttaac ggactaacct gagcttggcg gctctaaggt ctcttaagat tgtcagcttc 1860
agcctcaagc ttgaaggagg ctggagatgg agactggtct tgaaaagagt ccaggagagg 1920
cgggggctag gagacaaacg ggtgcaaacg tattggtggg agggacaccc agaaatgggt 1980
gccggggatg aggacacaaa ggtaaggcag gaggagctgc aggacccttg gaggaaatgg 2040
tgatcctcat ccccgggata cagggcaaag agtgtgtggt ggagaggaag tcggggacgt 2100
gccgagaagg cccaaaatag aactcggaat gtcgggggtc agaggcctag aggaaggagg 2160
actgtaagta agaggtgctt gggcatcctg gcagttgggg aggcctgaga tgggcacgcg 2220
gttgggtttg ggagcggagg agcaggctgc gggttaagaa gcccagacac aggtagaggg 2280
gagaacactt cgtcattttt gtggctaatc ggaagtggga gaccctcagg gaggttgaag 2340
atggaaaccc tacaaagggc ttcaacctga gaagcctgaa agacacgtat agctgagttg 2400
gacacaggcg acccgccagg gtttgaagca gcacccggcc tctagctgct atgccggaag 2460
tcgcaacacg caggtctcca gcctgcgagg aagggagagg tgggcgccag agctccagtc 2520
ccgcccctca ggactgtgcc tagcccacgg cgcagtagcc ccagccaagc taaattggac 2580
caatgggctc cggctcccgc ctcccgcccg ccgcggcctt ataaggagcc tcaagccggc 2640
tgtcggtctg gccccgcccc ctcgagtctc ttgtcctaat aaggacatct agggcatccc 2700
gcggctgggg gcgggcgaga gcgcgtctcc agggtaactg cgaagccccc ctctctctcc 2760
ccgttgctct tggaaactct aaggctcaga aaaggggggcg ctagcctcat attcatgtag 2820
ggtcgctcgc gggaggcggg gtcaagctgc tgtctccacc tggagcgccg ggagtccggg 2880
cgggggcggt gctgtggaat gtccccgact acaacccccg gtatgctctg caaggtaccc 2940
cactcccagg catcgctctc cgtttacacc cttaattccg tttataggat tagggccccc 3000
agttcccttt ctgggtttgc ctctgacttg tttgcgaccc agtcttttcc tgcagagtgt 3060
gaggctgtgg tcactgttct agaacaacaa aatcatcccg gagtcccccg ctcgtttctt 3120
ttttttttta aggcatatac taccacagaa cttgtctcag ttttgcggat gttagaataa 3180
ccatgtacct ggctatcaac tcgagagctg ggttcaagtc tggtttaata tctatttctt 3240
ctcattggca cgtaaaggca tgtaaagaag ccttgaattg aatgcacaaa ccctcgtctt 3300
taccctttaa cgcctcatga ctttgccaaa catctttaaa caagaacaaa gctctcagag 3360
atttccttgt acctttcccc tgtgcctttt cctcctattc ttgttttact gacgttggac 3420
atggcctttt ttcaggagac gagagtggct ttacacatgc agagtactag tgtttatccg 3480
actgacgacg ccttcatact tgtgatttcc tttgctttag gtcgggattc caggctaatc 3540
aatagttgct tcgattcaga aatgcaaaac cctaacctca ctcaaaaatt tcaggccaag 3600
ggatccagat aataccagat gggtcttaag aaagccgttt tgctgtkgga taaagagccg 3660
cttagtcggg gatcgttttc gggtcatttt actgagcgcc gcctcgccgg gctcagagyg 3720
gttcctggga aattgracca atgggctcgc gctgcgcgct gcggtgccgc ccaggacctg 3780
ggcctacatt tccctacrtg cactgcagct cggagccggc cggcgggaag actccgttac 3840
ccagygagcg aggcggcggc gcagggccag cggactccat ttcccgtcgg ctcgcggtgg 3900
gagcgccgga agcccgcccc acccctcatt gtgcggctcc tactaaacgg aaggggccgg 3960
gagaggccgc gttcagtcgg gtcccggcag cggctgcagc gctctcgtct tctgcggctc 4020

tcgstgccct ctccttttcg tttccggaaa catggtgagc ggcaggccac ggcgcctgag 4080
ggaggcggct gcgggtcggt cgcctgcgcg cgggaagcga ctggggagcg acgtccgcac 4140
cccctccccc agcgcccygg gcgggatgag ggtctcgcac gaaggggcgg gcggtgccgg 4200
aatgcgcgtg cgcgcccgtg ggcgccgggg agggccggtc tggacgtcgc tcgcgctccg 4260
cctgggctcc cctcccccac ccgctggtgc gcccgcgcgg acaccggccg cgggggggagg 4320
ggttcggggc gcgcgctggg gcgcccctcg cggaacggcc ggcgtcgcgc ctttcttctt 4380
aggggcgcgc tctcgactgg agcatgcggc gtccaaaccc ggcccggggg cggcgaggaa 4440
aaagcgcgcg agaggtctcg gcgcgcgccc cacccaaggc tgttcctgtc tgcgcgtcac 4500
cctccctcgg cagccgggtc cgtctcggga gcaggtggca gggatgcctg cgcccgggag 4560
gggcgggtgc tggcgcggcg aggggtgatg ctgggaagac ggggtcccgg ttggggcgtg 4620
ggtgcctgcg attaaggcct tgcaccgcaa ggcttggagg gggcatcgca gagaccgcgg 4680
cccgttcggg agcgcatcta acgagctgcg ttctcacccg tgcatcgggc gagggctgga 4740
acggcgctgt ctgtcggcgc gtgcgcgcac gctcaggccc ggccgccggt gccgaagccc 4800
tgggccagcg aggccttccc ggatgggcct gagtgagggt ggagccgagt ttagggaagt 4860
gacccaggcg ggcaccgccc ggaaaccacc ccgccccctg tatacggccc gagggctgag 4920
aaaaccgaag gttatgtagc ttgcccaggc tcctgcctgc aggagatggg tctgatacac 4980
ccgttgtttt agggcgattc cacagggtta gggacctgga gatgctgctg cctctcggat 5040
acgcgtttct gcattggtga gggggccggg cccagccgat ctcctgtccc cgcccactcc 5100
ggatgcggcc gagtcacgtg gccggcttct tctgcagttc cggggagttt gggggaccag 5160
atttaccttg gattgcccct ccctctcctg gctcgggcac cccgaagcac gacgcagagt 5220
aggaagagtt taaagacctc gaggcctctg ggcacttgag tttggcatgt taatttttat 5280
cagcgacttc tggggcctag caccattccc ggaagaaggg agttgtcggg cagggtcctt 5340
aatgggggtt gcaattcttg tcttggttgg gaaagagcct agctgggaac aggggtcgtt 5400
tgtgtagtaa ctgtattaag cagttctggt gggtttgttg taactgttcg gtgtccttgg 5460
ggtcccatct agcattgtgg ggtcttggct ctttctgaag aggagtcttg gctcaggtgt 5520
ttgggtcctc catatccgaa gttctaaatc accagtttta tccaaaggct gagagcaagt 5580
actgtgagta gcaaatgagt cagtagccgg tggatctttt aattctgggt tcacattttg 5640
gtcaaggttc tgcttttcac gagttttgca tttctatcct gtaaaatgag ggaaggggtt 5700
gttgaactag gtgatcagag gccatggctg ccattctgag gttgtgtccc atctggttct 5760
ttgacagtga cttgggtggg aaaactagtg acttgccttg ggaagggcac tagtgaccac 5820
tcgggacgtg gactgagatc ttctgaaaaa acaaggggca gttccctaga atccttttct 5880
ygacgtatac gtggcatgtc tcctaggcct ccggtgtggc tgtctctgat ggtgtcatca 5940
aggtgttcaa cgacatgaag gtgcgkaagt cttcaacgcc agaggaggtg aagaagcgca 6000
agaaggcggt gctcttctgc ctgagtgagg acaagaagaa catcatcctg gaggagggca 6060
aggagatcct ggtgggcgat gtgggccaga ctgtcgacga yccctacgcc acctttgtca 6120
agatgctgcc agataaggac tgccgctatg ccctctatga tgcaacctat gagaccaagg 6180
agagcaagaa ggaggatctg gtgtttatct tctggtgagc tcattctggc acttcttcct 6240
gtcacctgcc ccctttctca tgatgggaac ttytgtggct cctggtcaac ccagatgtgt 6300
gggtttggag ggggaggggg ttgtaacaaa agactcctca ctgccagccc aagggtttct 6360
caaagtcacc tgtggctttr ctgttgctgg gagttgcctg acgcgtgttt ctgtccaggg 6420
cccccgagtc tgcgcccctt aagagcaaaa tgatttatgc cagctccaag gacgccatca 6480
agaagaagct gacaggtaag ggccagcatt ggtgctgagt gcccttctgc tctggccttg 6540
gctgcagggc agggtgaatg gcaygcagag ggggtctgcc cccttgtctt cgctgcctgc 6600
ttgttccttc catctgcttt gattggctcc ttcccagcca cagaaactac ccccacccca 6660
gtgcttcctg ctcacagatg ctccccttte ttctttatag ggatcaagca tgaattgcaa 6720
gcaaactgct acgaggaggt caaggaccgc tgcaccctgg cagagaagct ggggggcagt 6780
gccgtcatct ccctggaggg caagcctttg tgagcccctt ctggcccccc gcctggagca 6840
tctggcagcc ccacacctgc ccttgggggt tgcaggctgc ccccttcctg ccagaccgga 6900

ggggctgggg ggatcccagc agggggaggg caatcccttc accccagttg ccaaacagac 6960
cccccacccc ctggattttc cttctccctc catcccttga cggttctggc cttcccaaac 7020
tgcttttgat cttttgattc ctcttgggct gaagcagacc aagttccccc caggcacccc 7080
agttgtgggg gagcctgtat ttttttaaac aacatcccca ttccccacct ggtcctcccc 7140
cttcccatgc tgccaacttc taaccgcaat agtgactctg tgcttgtctg tttagttctg 7200
tgtataaatg gaatgttgtg gagatgaccc ctccctgtgc cggctggttc ctctcccttt 7260
tcccctggtc acggctactc atggaagcag gaccagtaag ggaccttcga ttaaaaaaaa 7320
aaaagacaat aataaaaagg ctcattaatg ggatgtgttt ttcaaggttg ggacacagga 7380
gacttcagga ttgggggtgc actgggataa tggtgctatc tctctcctct gtactccaca 7440
cccaaccagc ccctgcacac ccttatctgg ctccaatgac tggctgagac gcctttctga 7500
tggtcaccat tattctgttg gtgctgtgat tgacactgga ctggctgttt aaggagtggg 7560
tgctggctac aggggtcttc cagcggaagt gatttatgcc tgaactgggt gctctgtagc 7620
cttgcctgga tggacagtga ggtccatctg atattctgtt ctgtcccttt ctaggtgtgg 7680
gctgataggc aagtcagagg tagctcaaag gttctgtttg gtctgctgtc ttccctggat 7740
ggctgactca agtttatgag ggtcaggttc agtggctgct gtgagtcagg ccaagttgag 7800
cctttgacct cctctagatg tggctggagt gcattcaggg tggggccagc caggtggagc 7860
ccgggacttc agaccccggt gtcttgagca ctgcctccct gtgtttaacc aaaagggtgt 7920
tgccggtacc ctgagccttt tgccagcatt gtcaagggct gtttccagcc tgtactttgg 7980
gctttttcct gcccagcagg caaagacata caagaaatct ggccctaggg tgaggtgaac 8040
tgggagtggt aatgtggcat ctatatttcc cttatcaact gaactgagga agagtcctcc 8100
tggtgttaca gttggctcca taactcctgg gaacttgtga gcggttctgg gaggaagggc 8160
tagatccctt ttctggttag gaaataggtg tggagaactt gggattcatt tttgtgctag 8220
ggtttatggg ggtaggggca aacctggaag acttctgggg ttcctaggtc ccagcttctg 8280
ggtgtcgcgc cttcccacag ggagcctgct gtcactggag agagggctga ggatgctgca 8340
ggggtgccct ccttggtgtc tccaggatct gatcatactg ccttggcaga aagtgtgcct 8400
ggtggacctt tgagctgtgg gaagggaagt aaggacaaca ggcctccttg ctacaaaacc 8460
agggttttta tttgcgaagg agaacaggcc gggatcaagt gagctatgat tccttcaggg 8520
ctatgggtga gggtagaggc tgttcctcag cttcctgggg agaggtgtca gggtccctgg 8580
ggctctgctg actgtggcca gttatgatgc ccatgctgcc ctgtgtcttg cagggcctgg 8640
cacctgtggt ggtggtgata gggggtttgg gctttccctg agaggatgtg gggcagagtg 8700
agggtgtggc tagtggcact gctagggaag gagacatctt gccggagaga ggggtctggg 8760
ataccctggt cactggagat cccagattag ggtctggctg agctctggct actctaaggc 8820
tccccttta gggcaacaag ggtgttccgg agaatcaggg tgctggcctg cagaggagaa 8880
acaggtcctt gaagtcccct gctccctgag tcaggctgcc attctttctg caccttagat 8940
gctccctcgc tccccacacc cctgctctct gcaataacag ctgctgttaa ttgggccttc 9000
ctacccagaa ggcacgggtc ctgagccagg gaagcatggc cccaagcctt agctggggct 9060
cctctcatgc acacacacgc gcatacatgc gtgcacaccc tcacatgcac gcacaccctg 9120
gtgcctgggg gagggaggga ggtgggaggt gggaggcttc acaagggccc agtgaggtgg 9180
gagggctcac cttggcgtgt ttgagctcca gctctgccag ctcaatgaga ttctttcgaa 9240
aagaggagac ccggcgggac ttgaagtcca tgagctctgg agaggagggg aagaggcatg 9300
gggatccggg cagggcaggg catgtggccc caggctgggg ctgggggctg cgggctcacc 9360
ttgcttggcg gagtcggaga ggcgctcgaa gcgttggcag cacagctgct ggtggctctc 9420
ggcgggccgc acctcccggt tcctggtgcg cgccttgtcc agcgccttgt tggcattctc 9480
gtagtcggcc agtgcccgca gccgccggta cagcaggtcc taggaccatg ggcacagggt 9540
cactgctgcc ctgtcctggc atccgggcct gagcgctgcg ctggggctgc ctctcacctt 9600
ggctgcctgt gagtcacgca tgtagtacct cagcatgtct gacagcttca ggtcctcatc 9660
ggaagccacc cggccctcca gcttctgcat cagaaaggtg ggcactggga agggagcttg 9720
tcagaggggc agctaagggg tgggaggcac ctcttcatct aaatatccag caaagaagcc 9780

ttaggtaact gttccaggtc aggtctggtg ttgaggggtg caggccaggg aaggcttcac 9840
agaggaagtg ggatctagtc tggatctcgg ggactgagta agagctggac agagggcgct 9900
gagggcttgg cagctgtgca agcccggcta tgtgttcagg gaaactttag ttcagggagg 9960
gcacttgggc aaagcagggg agggaggcag tggctggagc acaagaggct gaaagtggaa 10020
gctcaggact ttaggagctt ccagaagggg cagggtctct gaagggcctt gagcagggaa 10080
gagatgcggt ccaatagaca ttttaggaag atggcaacac tccagtgact ttcttaggga 10140
agagtgggga acatgtgtgc ccgggcagtg tgctggcagc tgggtgcagc ctgccagagg 10200
tttgggcaca cccttgtgcc tcccaaccct ccgctgctgc acacgtgcat acatgtgcat 10260
gcccctgcac cccccaacac acacacgaag ctggatcctt gtcttcaccc caccttttgt 10320
taaactagtc tggagtgtaa gcccaacttg ccactgggcc caagttatca ctgaacccaa 10380
gatgatcctg gagctcttcc ctctgccagg actcatcttc cctgcaaagc tcaccaggaa 10440
agcctttgag cacagaaggc agtactcacc ctcagccgtt caaagagctc tgccaatttg 10500
aggaagctcc taggaagcag tagggaggga gttagggtag ggggctgagg aacagctcaa 10560
agcctctctg tctcagcctc tccttgggaa tctccctctc ccctttcctt cccaacacca 10620
tcccccaaca ggagagccta cacctcccac ctagagctga cagtagacat ctacagagtg 10680
acccctcatg ggaagacatg gtattacagg cacagggaga gggtaggggt ggggggagtc 10740
ctcacgtctt tagctggttg acttcctgtg ttcccagact gctcagcgca gctgagatag 10800
ggatataatc gtc 10813

<210> 2
<211> 501
<212> DNA
<213> Homo sapiens

<400> 2
atggcctccg gtgtggctgt ctctgatggt gtcatcaagg tgttcaacga catgaaggtg 60
cgtaagtctt caacgccaga ggaggtgaag aagcgcaaga aggcggtgct cttctgcctg 120
agtgaggaca agaagaacat catcctggag gagggcaagg agatcctggt gggcgatgtg 180
ggccagactg tcgacgaccc ctacgccacc tttgtcaaga tgctgccaga taaggactgc 240
cgctatgccc tctatgatgc aacctatgag accaaggaga gcaagaagga ggatctggtg 300
tttatcttct gggcccccga gtctgcgccc cttaagagca aaatgattta tgccagctcc 360
aaggacgcca tcaagaagaa gctgacaggg atcaagcatg aattgcaagc aaactgctac 420
gaggaggtca aggaccgctg caccctggca gagaagctgg ggggcagtgc cgtcatctcc 480
ctggagggca agcctttgtg a 501

<210> 3
<211> 166
<212> PRT
<213> Homo sapiens

<400> 3
Met Ala Ser Gly Val Ala Val Ser Asp Gly Val Ile Lys Val Phe Asn
1 5 10 15

Asp Met Lys Val Arg Lys Ser Ser Thr Pro Glu Glu Val Lys Lys Arg
20 25 30

```
Lys Lys Ala Val Leu Phe Cys Leu Ser Glu Asp Lys Lys Asn Ile Ile
         35                  40                  45

Leu Glu Glu Gly Lys Glu Ile Leu Val Gly Asp Val Gly Gln Thr Val
     50                  55                  60

Asp Asp Pro Tyr Ala Thr Phe Val Lys Met Leu Pro Asp Lys Asp Cys
 65                  70                  75                  80

Arg Tyr Ala Leu Tyr Asp Ala Thr Tyr Glu Thr Lys Glu Ser Lys Lys
                 85                  90                  95

Glu Asp Leu Val Phe Ile Phe Trp Ala Pro Glu Ser Ala Pro Leu Lys
            100                 105                 110

Ser Lys Met Ile Tyr Ala Ser Ser Lys Asp Ala Ile Lys Lys Lys Leu
        115                 120                 125

Thr Gly Ile Lys His Glu Leu Gln Ala Asn Cys Tyr Glu Glu Val Lys
    130                 135                 140

Asp Arg Cys Thr Leu Ala Glu Lys Leu Gly Gly Ser Ala Val Ile Ser
145                 150                 155                 160

Leu Glu Gly Lys Pro Leu
                165
```

<210> 4
<211> 15
<212> DNA
<213> Homo sapiens

<400> 4
ttgctgtkgg ataaa 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
ctcagagygg ttcct 15

<210> 6

<211> 15
<212> DNA
<213> Homo sapiens

<400> 6
gaaattgrac caatg 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
tccctacrtg cactg 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
tacccagyga gcgag 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
gctctcgstg ccctc 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
agcgcccygg gcggg 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
cttttctyga cgtat 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
aggtgcgkaa gtctt 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapiens

<400> 13
ggaacttytg tggct 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapiens

<400> 14
tggctttrct gttgc 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
aatggcaygc agagg 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

<400> 16
gccgttttgc tgtkg 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
cggctcttta tccma 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
gccgggctca gagyg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19
tttcccagga accrc 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
tcctgggaaa ttgra 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
cgagcccatt ggtyc 15

<210> 22

<211> 15
<212> DNA
<213> Homo sapiens

<400> 22
tacatttccc tacrt 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapiens

<400> 23
gagctgcagt gcayg 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
ctccgttacc cagyg 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
cgccgcctcg ctcrc 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
tctgcggctc tcgst 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
aaaggagagg gcasc 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
tcccccagcg cccyg 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
cctcatcccg cccrg 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapiens

<400> 30
agaatccttt tctyg 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
ccacgtatac gtcra 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapiens

<400> 32
acatgaaggt gcgka 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
gcgttgaaga cttmc 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapiens

<400> 34
atgatgggaa cttyt 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapiens

<400> 35
accaggagcc acara 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapiens

<400> 36
cacctgtggc tttrc 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapiens

<400> 37
ctcccagcaa cagya 15

<210> 38

<211> 15
<212> DNA
<213> Homo sapiens

<400> 38
agggtgaatg gcayg 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapiens

<400> 39
agacccccctc tgcrt 15

<210> 40
<211> 10
<212> DNA
<213> Homo sapiens

<400> 40
gttttgctgt 10

<210> 41
<211> 10
<212> DNA
<213> Homo sapiens

<400> 41
ctctttatcc 10

<210> 42
<211> 10
<212> DNA
<213> Homo sapiens

<400> 42
gggctcagag 10

<210> 43
<211> 10
<212> DNA
<213> Homo sapiens

<400> 43
cccaggaacc 10

<210> 44
<211> 10
<212> DNA
<213> Homo sapiens

<400> 44
tgggaaattg 10

<210> 45
<211> 10
<212> DNA
<213> Homo sapiens

<400> 45
gcccattggt 10

<210> 46
<211> 10
<212> DNA
<213> Homo sapiens

<400> 46
atttccctac 10

<210> 47
<211> 10
<212> DNA
<213> Homo sapiens

<400> 47
ctgcagtgca 10

<210> 48
<211> 10
<212> DNA
<213> Homo sapiens

<400> 48
cgttacccag 10

<210> 49
<211> 10
<212> DNA
<213> Homo sapiens

<400> 49
cgcctcgctc 10

<210> 50
<211> 10
<212> DNA
<213> Homo sapiens

<400> 50
gcggctctcg 10

<210> 51
<211> 10
<212> DNA
<213> Homo sapiens

<400> 51
ggagagggca 10

<210> 52
<211> 10
<212> DNA
<213> Homo sapiens

<400> 52
cccagcgccc 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapiens

<400> 53
catcccgccc 10

<210> 54

<211> 10
<212> DNA
<213> Homo sapiens

<400> 54
atccttttct 10

<210> 55
<211> 10
<212> DNA
<213> Homo sapiens

<400> 55
cgtatacgtc 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapiens

<400> 56
tgaaggtgcg 10

<210> 57
<211> 10
<212> DNA
<213> Homo sapiens

<400> 57
ttgaagactt 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapiens

<400> 58
atgggaactt 10

<210> 59
<211> 10
<212> DNA
<213> Homo sapiens

<400> 59
aggagccaca 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapiens

<400> 60
ctgtggcttt 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapiens

<400> 61
ccagcaacag 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapiens

<400> 62
gtgaatggca 10

<210> 63
<211> 10
<212> DNA
<213> Homo sapiens

<400> 63
ccccctctgc 10

<210> 64
<211> 10813
<212> DNA
<213> Homo sapiens

<220>
<221> allele

<222> (3647)
<223> PS1: Polymorphic base G or T

<220>
<221> allele
<222> (3719)
<223> PS2: Polymorphic base C or T

<220>
<221> allele
<222> (3736)
<223> PS3: Polymorphic base G or A

<220>
<221> allele
<222> (3798)
<223> PS4: Polymorphic base A or G

<220>
<221> allele
<222> (3845)
<223> PS5: Polymorphic base C or T

<220>
<221> allele
<222> (4024)
<223> PS6: Polymorphic base G or C

<220>
<221> allele
<222> (4158)
<223> PS7: Polymorphic base C or T

<220>
<221> allele
<222> (5881)
<223> PS8: Polymorphic base T or C

<220>
<221> allele
<222> (5966)
<223> PS9: Polymorphic base T or G

<220>
<221> allele
<222> (6101)
<223> PS10: Polymorphic base C or T

<220>
<221> allele
<222> (6273)
<223> PS11: Polymorphic base C or T

<220>
<221> allele
<222> (6380)
<223> PS12: Polymorphic base G or A

<220>
<221> allele
<222> (6564)
<223> PS13: Polymorphic base C or T

<400> 64
caggtggctg ggataatctt tgcctgccct gcctgggcca ggcccctcgg ggaacctccc 60
accccactgc catggttccg cctcccggtg tagcccagtg ctcaccaggt gctcccggta 120
gagcaccagc agtatcactc gggctcctga gtgccgagct ggccagtagc tgccagagcc 180
tcccgctttg ggctgggcag gaacctgaaa ggtacagagg agttctgaac atgaaagcca 240
gccagtcatc tttgaataag tccccagcag gctttgtggc cctgggtaag tcactcagtt 300
gtctctcagt ttccttctct gtaaaatggg ccaggtgatg ggatgatccg ctgcaaatct 360
gggataaaac agcaggcagt ttcggggcag tgggtggggg tgtctgccca gcctggcctc 420
acttctacca tgcgctgcag gacaggtgta attagacagg aaaactcaat aatggtgacc 480
ccgtaactaa ctgagctggg atccttaagt agcttgatca tggaattctc tcattggctg 540
ttaccaatat ttctattcac aggttaaaat actaagtcga aaagaggata agtcatttgc 600
tcaaggccag gcatcactgg tgagatttgg actcaggctt gccaaatccc ccaagccaga 660
aaatcatact cccgcactcc cactttgcct tccactcccc ttgccccttc ctcactggca 720
tggaagagtc ctggacttcc gcaagtcgcc cctgcggggc tggactgttc tctccagatg 780
gtgagtccgg ggcatggtca cctgcgggag agaaaagagt taactccaca ctccacgaat 840
ccggcagctg cgaggatctg ggatgcaacc tactcagatg gttacgtgag accggtgagg 900
gaggctcagg gagcgccagc cactagccct aggccgaacg gcaactggga caggagtggg 960
gagcttaaaa gccaagggct cccaggagag tacggaggct ccacctgggg gaggtggggg 1020
atcaccgacc cctttccgag gcgctcaccg cccgatgttc ggtgccgctg cagccgctcg 1080
tccagcatcc ggcagagccc gtctccgaag tgctgtagga tcttagcttc cttcccgctg 1140
cgcagcggca gtgggtaccg tcggagggaa cgcagcgcct ggccgggcag gggtagggta 1200
cctggtcagg gcaggcctga cctggccagc agcttttccc atcgggccac gccaggaccc 1260
accttctgaa atacgaagcg cgtgcggcgc ctgctgcggg tcgcctcgtc ccgccactcg 1320
gtcagccagc gaacgaagag cgggttggga caggcaggca gcgggcgctt ccggcccagg 1380
cggaccgggg ccgccatgag cccgagggcg ggtcctccgg cgctccacgc ccgcgggact 1440
gggacgccgc cagttctgga gtcgggattc gaacacctgg cccagggcgg ggcgggggaa 1500
gagggagacc gagggcagga ccgttgagat cacggggccc caaacacagg gggcactaac 1560
gagaggagag cctgcggccg ccctcccacc cctgcgctgc tctaagtggt tggtcagggg 1620
cgctgtctct aacgatcttt tcccttggct ccactccagc cagcctttga gacgcgccct 1680
ttcccagccc ccggagcctg gaccggctcc ttcctgtcac cccagattcg agagcggggt 1740
ctttgtgtag gcgagaggaa ggcggacagc gccccctgtg acggggagga ctggctgggc 1800
ggcgcttaac ggactaacct gagcttggcg gctctaaggt ctcttaagat tgtcagcttc 1860
agcctcaagc ttgaaggagg ctggagatgg agactggtct tgaaaagagt ccaggagagg 1920

cgggggctag gagacaaacg ggtgcaaacg tattggtggg agggacaccc agaaatgggt 1980
gccggggatg aggacacaaa ggtaaggcag gaggagctgc aggacccttg gaggaaatgg 2040
tgatcctcat ccccgggata cagggcaaag agtgtgtggt ggagaggaag tcggggacgt 2100
gccgagaagg cccaaaatag aactcggaat gtcgggggtc agaggcctag aggaaggagg 2160
actgtaagta agaggtgctt gggcatcctg gcagttgggg aggcctgaga tgggcacgcg 2220
gttgggtttg ggagcggagg agcaggctgc gggttaagaa gcccagacac aggtagaggg 2280
gagaacactt cgtcattttt gtggctaatc ggaagtggga gaccctcagg gaggttgaag 2340
atggaaaccc tacaaagggc ttcaacctga gaagcctgaa agacacgtat agctgagttg 2400
gacacaggcg acccgccagg gtttgaagca gcacccggcc tctagctgct atgccggaag 2460
tcgcaacacg caggtctcca gcctgcgagg aagggagagg tgggcgccag agctccagtc 2520
ccgcccctca ggactgtgcc tagcccacgg cgcagtagcc ccagccaagc taaattggac 2580
caatgggctc cggctcccgc ctcccgcccg ccgcggcctt ataaggagcc tcaagccggc 2640
tgtcggtctg gccccgcccc ctcgagtctc ttgtcctaat aaggacatct agggcatccc 2700
gcggctgggg gcgggcgaga gcgcgtctcc agggtaactg cgaagccccc ctctctctcc 2760
ccgttgctct tggaaactct aaggctcaga aaaggggcgg ctagcctcat attcatgtag 2820
ggtcgctcgc gggaggcggg gtcaagctgc tgtctccacc tggagcgccg ggagtccggg 2880
cgggggcggt gctgtggaat gtccccgact acaacccccg gtatgctctg caaggtaccc 2940
cactcccagg catcgctctc cgtttacacc cttaattccg tttataggat tagggccccc 3000
agttcccttt ctgggtttgc ctctgacttg tttgcgaccc agtcttttcc tgcagagtgt 3060
gaggctgtgg tcactgttct agaacaacaa aatcatcccg gagtcccccg ctcgtttctt 3120
ttttttttta aggcatatac taccacagaa cttgtctcag ttttgcggat gttagaataa 3180
ccatgtacct ggctatcaac tcgagagctg ggttcaagtc tggtttaata tctatttctt 3240
ctcattggca cgtaaaggca tgtaaagaag ccttgaattg aatgcacaaa ccctcgtctt 3300
tacccttaa cgcctcatga ctttgccaaa catctttaaa caagaacaaa gctctcagag 3360
atttccttgt acctttcccc tgtgcctttt cctcctattc ttgttttact gacgttggac 3420
atggcctttt ttcaggagac gagagtggct ttacacatgc agagtactag tgtttatccg 3480
actgacgacg ccttcatact tgtgatttcc tttgctttag gtcgggattc caggctaatc 3540
aatagttgct tcgattcaga aatgcaaaac cctaacctca ctcaaaaatt tcaggccaag 3600
ggatccagat aataccagat gggtcttaag aaagccgttt tgctgtkgga taaagagccg 3660
cttagtcggg gatcgttttc gggtcatttt actgagcgcc gcctcgccgg gctcagagyg 3720
gttcctggga aattgracca atgggctcgc gctgcgcgct gcggtgccgc ccaggacctg 3780
ggcctacatt tccctacrtg cactgcagct cggagccggc cggcgggaag actccgttac 3840
ccagygagcg aggcggcggc gcagggccag cggactccat ttcccgtcgg ctcgcggtgg 3900
gagcgccgga agcccgcccc acccctcatt gtgcggctcc tactaaacgg aaggggccgg 3960
gagaggccgc gttcagtcgg gtcccggcag cggctgcagc gctctcgtct tctgcggctc 4020
tcgstgccct ctccttttcg tttccggaaa catggtgagc ggcaggccac ggcgcctgag 4080
ggaggcggct gcgggtcggt cgcctgcgcg cgggaagcga ctggggagcg acgtccgcac 4140
cccctccccc agcgcccygg gcgggatgag ggtctcgcac gaaggggcgg gcggtgccgg 4200
aatgcgcgtg cgcgcccgtg ggcgccgggg agggccggtc tggacgtcgc tcgcgctccg 4260
cctgggctcc cctcccccac ccgctggtgc gcccgcgcgg acaccggccg cgggggggagg 4320
ggttcggggc gcgcgctggg gcgcccctcg cggaacggcc ggcgtcgcgc ctttcttctt 4380
aggggcgcgc tctcgactgg agcatgcggc gtccaaaccc ggcccggggg cggcgaggaa 4440
aaagcgcgcg agaggtctcg gcgcgcgccc cacccaaggc tgttcctgtc tgcgcgtcac 4500
cctccctcgg cagccgggtc cgtctcggga gcaggtggca gggatgcctg cgcccgggag 4560
gggcgggtgc tggcgcggcg aggggtgatg ctgggaagac ggggtcccgg ttggggcgtg 4620
ggtgcctgcg attaaggcct tgcaccgcaa ggcttggagg gggcatcgca gagaccgcgg 4680
cccgttcggg agcgcatcta acgagctgcg ttctcacccg tgcatcgggc gagggctgga 4740
acggcgctgt ctgtcggcgc gtgcgcgcac gctcaggccc ggccgccggt gccgaagccc 4800

tgggccagcg aggccttccc ggatgggcct gagtgagggt ggagccgagt ttagggaagt 4860
gacccaggcg ggcaccgccc ggaaaccacc ccgccccctg tatacggccc gagggctgag 4920
aaaaccgaag gttatgtagc ttgcccaggc tcctgcctgc aggagatggg tctgatacac 4980
ccgttgtttt agggcgattc cacagggtta gggacctgga gatgctgctg cctctcggat 5040
acgcgtttct gcattggtga gggggccggg cccagccgat ctcctgtccc cgcccactcc 5100
ggatgcggcc gagtcacgtg gccggcttct tctgcagttc cggggagttt gggggaccag 5160
atttaccttg gattgcccct ccctctcctg gctcgggcac cccgaagcac gacgcagagt 5220
aggaagagtt taaagacctc gaggcctctg ggcacttgag tttggcatgt taatttttat 5280
cagcgacttc tggggcctag caccattccc ggaagaaggg agttgtcggg cagggtcctt 5340
aatgggggtt gcaattcttg tcttggttgg gaaagagcct agctgggaac aggggtcgtt 5400
tgtgtagtaa ctgtattaag cagttctggt gggtttgttg taactgttcg gtgtccttgg 5460
ggtcccatct agcattgtgg ggtcttggct ctttctgaag aggagtcttg gctcaggtgt 5520
ttgggtcctc catatccgaa gttctaaatc accagtttta tccaaaggct gagagcaagt 5580
actgtgagta gcaaatgagt cagtagccgg tggatctttt aattctgggt tcacattttg 5640
gtcaaggttc tgcttttcac gagttttgca tttctatcct gtaaaatgag ggaaggggtt 5700
gttgaactag gtgatcagag gccatggctg ccattctgag gttgtgtccc atctggttct 5760
ttgacagtga cttgggtggg aaaactagtg acttgccttg ggaagggcac tagtgaccac 5820
tcgggacgtg gactgagatc ttctgaaaaa acaaggggca gttccctaga atccttttct 5880
ygacgtatac gtggcatgtc tcctaggcct ccggtgtggc tgtctctgat ggtgtcatca 5940
aggtgttcaa cgacatgaag gtgcgkaagt cttcaacgcc agaggaggtg aagaagcgca 6000
agaaggcggt gctcttctgc ctgagtgagg acaagaagaa catcatcctg gaggagggca 6060
aggagatcct ggtgggcgat gtgggccaga ctgtcgacga yccctacgcc acctttgtca 6120
agatgctgcc agataaggac tgccgctatg ccctctatga tgcaacctat gagaccaagg 6180
agagcaagaa ggaggatctg gtgtttatct tctggtgagc tcattctggc acttcttcct 6240
gtcacctgcc ccctttctca tgatgggaac ttytgtggct cctggtcaac ccagatgtgt 6300
gggtttggag ggggaggggg ttgtaacaaa agactcctca ctgccagccc aagggtttct 6360
caaagtcacc tgtggctttr ctgttgctgg gagttgcctg acgcgtgttt ctgtccaggg 6420
cccccgagtc tgcgcccctt aagagcaaaa tgatttatgc cagctccaag gacgccatca 6480
agaagaagct gacaggtaag ggccagcatt ggtgctgagt gcccttctgc tctggccttg 6540
gctgcagggc agggtgaatg gcaygcagag ggggtctgcc cccttgtctt cgctgcctgc 6600
ttgttccttc catctgcttt gattggctcc ttcccagcca cagaaactac ccccacccca 6660
gtgcttcctg ctcacagatg ctccccttte ttctttatag ggatcaagca tgaattgcaa 6720
gcaaactgct acgaggaggt caaggaccgc tgcaccctgg cagagaagct ggggggcagt 6780
gccgtcatct ccctggaggg caagcctttg tgagcccctt ctggcccccc gcctggagca 6840
tctggcagcc ccacacctgc ccttgggggt tgcaggctgc ccccttcctg ccagaccgga 6900
ggggctgggg ggatcccagc agggggaggg caatcccttc accccagttg ccaaacagac 6960
cccccacccc ctggattttc cttctccctc catcccttga cggttctggc cttcccaaac 7020
tgcttttgat cttttgattc ctcttgggct gaagcagacc aagttccccc caggcacccc 7080
agttgtgggg gagcctgtat ttttttaac aacatcccca ttccccacct ggtcctcccc 7140
cttcccatgc tgccaacttc taaccgcaat agtgactctg tgcttgtctg tttagttctg 7200
tgtataaatg gaatgttgtg gagatgaccc ctccctgtgc cggctggttc ctctcccttt 7260
tcccctggtc acggctactc atggaagcag gaccagtaag ggaccttcga ttaaaaaaaa 7320
aaaagacaat aataaaaagg ctcattaatg ggatgtgttt ttcaaggttg ggacacagga 7380
gacttcagga ttgggggtgc actgggataa tggtgctatc tctctcctct gtactccaca 7440
cccaaccagc ccctgcacac ccttatctgg ctccaatgac tggctgagac gcctttctga 7500
tggtcaccat tattctgttg gtgctgtgat tgacactgga ctggctgttt aaggagtggg 7560
tgctggctac aggggtcttc cagcggaagt gatttatgcc tgaactgggt gctctgtagc 7620
cttgcctgga tggacagtga ggtccatctg atattctgtt ctgtcccttt ctaggtgtgg 7680

gctgataggc aagtcagagg tagctcaaag gttctgtttg gtctgctgtc ttccctggat 7740
ggctgactca agtttatgag ggtcaggttc agtggctgct gtgagtcagg ccaagttgag 7800
cctttgacct cctctagatg tggctggagt gcattcaggg tggggccagc caggtggagc 7860
ccgggacttc agaccccggt gtcttgagca ctgcctccct gtgtttaacc aaaagggtgt 7920
tgccggtacc ctgagccttt tgccagcatt gtcaagggct gtttccagcc tgtactttgg 7980
gctttttcct gcccagcagg caaagacata caagaaatct ggccctaggg tgaggtgaac 8040
tgggagtggt aatgtggcat ctatatttcc cttatcaact gaactgagga agagtcctcc 8100
tggtgttaca gttggctcca taactcctgg gaacttgtga gcggttctgg gaggaagggc 8160
tagatccctt ttctggttag gaaataggtg tggagaactt gggattcatt tttgtgctag 8220
ggtttatggg ggtaggggca aacctggaag acttctgggg ttcctaggtc ccagcttctg 8280
ggtgtcgcgc cttcccacag ggagcctgct gtcactggag agagggctga ggatgctgca 8340
ggggtgccct ccttggtgtc tccaggatct gatcatactg ccttggcaga aagtgtgcct 8400
ggtggacctt tgagctgtgg gaagggaagt aaggacaaca ggcctccttg ctacaaaacc 8460
agggttttta tttgcgaagg agaacaggcc gggatcaagt gagctatgat tccttcaggg 8520
ctatgggtga gggtagaggc tgttcctcag cttcctgggg agaggtgtca gggtccctgg 8580
ggctctgctg actgtggcca gttatgatgc ccatgctgcc ctgtgtcttg cagggcctgg 8640
cacctgtggt ggtggtgata gggggtttgg gctttccctg agaggatgtg gggcagagtg 8700
agggtgtggc tagtggcact gctagggaag gagacatctt gccggagaga ggggtctggg 8760
ataccctggt cactggagat cccagattag ggtctggctg agctctggct actctaaggc 8820
tcccccttta gggcaacaag ggtgttccgg agaatcaggg tgctggcctg cagaggagaa 8880
acaggtcctt gaagtcccct gctccctgag tcaggctgcc attctttctg caccttagat 8940
gctccctcgc tccccacacc cctgctctct gcaataacag ctgctgttaa ttgggccttc 9000
ctacccagaa ggcacgggtc ctgagccagg gaagcatggc cccaagcctt agctggggct 9060
cctctcatgc acacacacgc gcatacatgc gtgcacaccc tcacatgcac gcacacctg 9120
gtgcctgggg gagggaggga ggtgggaggt gggaggcttc acaagggccc agtgaggtgg 9180
gagggctcac cttggcgtgt ttgagctcca gctctgccag ctcaatgaga ttctttcgaa 9240
aagaggagac ccggcgggac ttgaagtcca tgagctctgg agaggagggg aagaggcatg 9300
gggatccggg cagggcaggg catgtggccc caggctgggg ctgggggctg cgggctcacc 9360
ttgcttggcg gagtcggaga ggcgctcgaa gcgttggcag cacagctgct ggtggctctc 9420
ggcgggccgc acctcccggt tcctggtgcg cgccttgtcc agcgccttgt tggcattctc 9480
gtagtcggcc agtgcccgca gccgccggta cagcaggtcc taggaccatg ggcacagggt 9540
cactgctgcc ctgtcctggc atccgggcct gagcgctgcg ctggggctgc ctctcacctt 9600
ggctgcctgt gagtcacgca tgtagtacct cagcatgtct gacagcttca ggtcctcatc 9660
ggaagccacc cggccctcca gcttctgcat cagaaaggtg ggcactggga agggagcttg 9720
tcagaggggc agctaagggg tgggaggcac ctcttcatct aaatatccag caaagaagcc 9780
ttaggtaact gttccaggtc aggtctggtg ttgaggggtg caggccaggg aaggcttcac 9840
agaggaagtg ggatctagtc tggatctcgg ggactgagta agagctggac agagggcgct 9900
gagggcttgg cagctgtgca agcccggcta tgtgttcagg gaaactttag ttcagggagg 9960
gcacttgggc aaagcagggg agggaggcag tggctggagc acaagaggct gaaagtggaa 10020
gctcaggact ttaggagctt ccagaagggg cagggtctct gaagggcctt gagcagggaa 10080
gagatgcggt ccaatagaca ttttaggaag atggcaacac tccagtgact ttcttaggga 10140
agagtgggga acatgtgtgc ccgggcagtg tgctggcagc tgggtgcagc ctgccagagg 10200
tttgggcaca cccttgtgcc tcccaaccct ccgctgctgc acacgtgcat acatgtgcat 10260
gcccctgcac cccccaacac acacacgaag ctggatcctt gtcttcaccc caccttttgt 10320
taaactagtc tggagtgtaa gcccaacttg ccactgggcc caagttatca ctgaacccaa 10380
gatgatcctg gagctcttcc ctctgccagg actcatcttc cctgcaaagc tcaccaggaa 10440
agcctttgag cacagaaggc agtactcacc ctcagccgtt caaagagctc tgccaatttg 10500
aggaagctcc taggaagcag tagggaggga gttagggtag ggggctgagg aacagctcaa 10560

agcctctctg tctcagcctc tccttgggaa tctccctctc ccctttcctt cccaacacca 10620
tcccccaaca ggagagccta cacctcccac ctagagctga cagtagacat ctacagagtg 10680
tcccctcatg ggaagacatg gtattacagg cacagggaga gggtaggggt gggggagtc 10740
ttcacgtctt tagctggttg acttcctgtg ttcccagact gctcagcgca gctgagatag 10800
ggatataatc gtc 10813

<210> 65
<211> 1500
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (30)
<223> PS1: Polymorphic base G or T

<220>
<221> misc_feature
<222> (61)..(120)
<223> Nucleotides represent sequence between PS1 and PS2

<220>
<221> allele
<222> (150)
<223> PS2: Polymorphic base C or T

<220>
<221> misc_feature
<222> (181)..(240)
<223> Nucleotides represent sequence between PS2 and PS3

<220>
<221> allele
<222> (270)
<223> PS3: Polymorphic base G or A

<220>
<221> misc_feature
<222> (301)..(360)
<223> Nucleotides represent sequence between PS3 and PS4

<220>
<221> allele
<222> (390)
<223> PS4: Polymorphic base A or G

<220>

<221> misc_feature
<222> (421)..(480)
<223> Nucleotides represent sequence between PS4 and PS5

<220>
<221> allele
<222> (510)
<223> PS5: Polymorphic base C or T

<220>
<221> misc_feature
<222> (541)..(600)
<223> Nucleotides represent sequence between PS5 and PS6

<220>
<221> allele
<222> (630)
<223> PS6: Polymorphic base G or C

<220>
<221> misc_feature
<222> (661)..(720)
<223> Nucleotides represent sequence between PS6 and PS7

<220>
<221> allele
<222> (750)
<223> PS7: Polymorphic base C or T

<220>
<221> misc_feature
<222> (781)..(840)
<223> Nucleotides represent sequence between PS7 and PS8

<220>
<221> allele
<222> (870)
<223> PS8: Polymorphic base T or C

<220>
<221> misc_feature
<222> (901)..(960)
<223> Nucleotides represent sequence between PS8 and PS9

<220>
<221> allele
<222> (990)
<223> PS9: Polymorphic base T or G

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> Nucleotides represent sequence between PS9 and PS10

<220>
<221> allele
<222> (1110)
<223> PS10: Polymorphic base T or C

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> Nucleotides represent sequence between PS10 and PS11

<220>
<221> allele
<222> (1230)
<223> PS11: Polymorphic base T or C

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> Nucleotides represent sequence between PS11 and PS12

<220>
<221> allele
<222> (1350)
<223> PS12: Polymorphic base G or A

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> Nucleotides represent sequence between PS12 and PS13

<220>
<221> allele
<222> (1470)
<223> PS13: Polymorphic base C or T

<400> 65
gatgggtctt aagaaagccg ttttgctgtk ggataaagag ccgcttagtc ggggatcgtt 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

tactgagcgc cgcctcgccg ggctcagagy ggttcctggg aaattggcca atgggctcag 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
gccgggctca gagcgttcct gggaaattgr accaatgggc tcgcgctgcg cgctgcggtg 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
gcccaggacc tgggcctaca tttccctacr tgcactgcag ctcggagccg gccggcggga 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
ccggccggcg ggaagactcc gttacccagy gagcgaggcg gcggcgcagg gccagcggac 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
tgcagcgctc tcgtcttctg cggctctcgs tgccctctcc ttttcgtttc cggaaacatg 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
cgacgtccgc acccccctcc ccagcgcccy gggcgggatg agggtctcgc acgaaggggc 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
caaggggcag ttccctagaa tcctttttcty gacgtatacg tggcatgtct cctaggcctc 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
atcaaggtgt tcaacgacat gaaggtgcgk aagtcttcaa cgccagagga ggtgaagaag 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
gtgggcgatg tgggccagac tgtcgacgay ccctacgcca cctttgtcaa gatgctgcca 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
acctgccccc tttctcatga tgggaactty tgtggctcct ggtcaaccca gatgtgtggg 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
aagggtttct caaagtcacc tgtggctttr ctgttgctgg gagttgcctg acgcgtgttt 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
gccttggctg cagggcaggg tgaatggcay gcagaggggg tctgcccccc tgtcttcgct 1500

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/18815

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/041; C12Q 1/68
US CL : 536/23.1; 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1; 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ONO et al. "Characterization of a novel cofilin isoform that is predominatly expressed in mammalian skeletal muscle". J. Biol. chem. May 1994, Vol 269, No. 21, pages 15280-15286.	1-2
X	OGAWA et al. "Coding sequence of human placenta cofilin cDNA". Nucleic Acids Research. October 1990, Vol. 18, No. 23, pages7169.	1-2

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 October 2001 (10.10.2001)	16 NOV 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Jeanine Enewold Goldberg Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/18815

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-2, in part

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/18815

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Groups 1-13, claim(s) 1-2 in part, drawn to methods for haplotyping CFL1 comprising determining which of the 17 haplotypes listed in the table in claim 1 defines one copy of the individual's CFL1 gene. It is noted that Groups 1-17 correspond to the 17 haplotypes listed in the claim. Therefore, the first mentioned invention is the methods of claims 1-2 to the extent that they apply to haplotype 1. **Group 1, the first mentioned invention, is the invention which will be searched in accordance with PCT Article 17(3)(a).** Additional groups may be elected. For example, if Group 2 is elected, and the proper fees are paid, then claims 1-2 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the second haplotype of Table 4 of the CFL1 gene. Upon election of an invention to be searched in addition to group 1, please identify the number of the haplotype or haplotype pair to be searched.

Groups 14-36, claim(s) 3-4 in part, drawn to methods of haplotyping the CFL1 comprising determining which of the 21 haplotype pairs listed in the table in claim 3 defines both copies of an individual's CFL1 gene. Groups 14-36 correspond the 22 haplotype pairs listed in the table in claim 3. If group 14 is elected, then claims 3-4 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the first listed haplotype pair. Upon election of one or more inventions in this group, please specify the haplotype pairs requested for search.

Groups 36-48, claim(s) 5-7, in part drawn to a method for genotyping the CFL1 gene It is noted that Groups 36-48 correspond to polymorphic sites PS1-9, PS 11-13, respectively. For example, if Group 36 is elected, claims 5-7 will be examined to the extent that they apply are limited to method of genotyping comprising a step of identifying the nucleotide pair at PS1.

Groups 49-214, claim(s) 8-9 in part drawn to a method for haplotyping the CFL1 gene by identifying a CFL1 genotype for the individual at two or more polymorphic sites PS1-9, PS 11-13. It is noted that the claims encompass methods requiring identification of 166 possible combinations of two of the recited polymorphic sites, and that Groups 49-214 each correspond to one of these possible pairs, in the order recited in the claim. For example, if Group 49 is selected, then claim 8 will be examined to the extent that it applies to a combination of PS1 and PS2. If Group 214 is selected, then claim 8 will be examined to the extent that it applies to a combination of PS12 and PS13. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype pair.

Groups 215-380, claim(s) 10-11, in part drawn to a method for predicting a haplotype pair for the CFL1 gene by identifying a CFL1 genotype for the individual at two or more polymorphic sites PS1-9, PS 11-13. It is noted that the claims encompass methods requiring identification of 105 possible combinations of two of the recited polymorphic sites, and that Groups 215-380 each correspond to one of these possible pairs, in the order recited in the claim. For example, if Group 215 is selected, then claim 10 will be examined to the extent that it applies to a combination of PS1 and PS2. If Group 380 is selected, then claim 10 will be examined to the extent that it applies to a combination of PS12 and PS13. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype pair.

Groups 381-417, claim(s) 12-13, in part drawn to a method for identifying an association between a trait and a haplotype between one of the 13 haplotypes and 22 haplotype pairs of CFL1 gene. Groups 381-417 each correspond to one of the particular combinations of the polymorphic sites, haplotypes, and the haplotype pairs encompassed by the claims (i.e., the 13 different haplotypes, as well as the 22 different haplotype pairs). For example if Group 381 is selected, the claims will be examined to the extent that they apply to the first haplotype.

Groups 418-430, claim(s) 14-19, in part, drawn to a composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the CFL1 gene.

Group 431432, claim 20-21, drawn to a kit comprising a set of oligonucleotides designed to genotype each of the polymorphic sites.

Groups 433-459, claims 22-23 and 26-27, in part, drawn to a polynucleotide which is a polymorphic variant of a reference sequence for CFL1 gene or a fragment thereof. Claims 22-23 and 26-27 recite 15 different isogenes and 12 fragments comprising polymorphisms.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/18815

Group 460-474, claim(s) 24-25 and 28-29, in part drawn to a recombinant nonhuman organisms comprising a polynucleotide which is a polymorphic variant of a reference sequence for CFL1 gene.

Group 475-511, claim(s) 30, drawn to a computer system comprising polymorphism data wherein the data comprises the haplotypes and haplotype pairs listed in the claims.

Groups 512-526, claim(s) 31, in part, drawn to genome anthologies comprising any one CFL1 isogenes having any one of the haplotype set forth in the table. It is noted that Groups 512-526 correspond to anthologies comprising one of the haplotypes 1-15 listed in the table in claim 31 in the order shown in table. For example, Group 512 is drawn to an anthology comprising haplotype 1.

The inventions listed in the instant application lack unity for a number of reasons.

The first claimed inventions, claims 1-2 (groups 1-15) lack unity because they represent methods which have different results depending on the nucleic acid present in the sample. That is, depending on the nucleic acid present in the sample, the special technical feature of each part of the invention would be the haplotype listed in the table. Since these methods result in different outcomes, they lack unity with one another.

Each polymorphic site and each molecule containing said polymorphic site is structurally and functionally distinct from and has a different special technical feature than each other polymorphic site and molecules containing said site. The chemical structure of each polymorphism and of each molecule containing the same differ from each other. For example, a polynucleotide comprising PS1 is chemically, structurally, and functionally different from a molecule comprising PS4. As the products and methods encompassed by the claims do not share a special technical feature, the distinct products and methods may not properly be presented in the alternative. Accordingly, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed by the claims, and the claims will be examined only as they read upon the invention of the elected group. For the same reasons, the remainder of the claims have been separated in a number of groups corresponding to the number of different inventions encompassed thereby.

The haplotypes and genotypes encompassed by the instantly recited method claims are also distinct from each other and from the single polymorphisms recited in e.g., claims 5-6. For example, a molecule of haplotype 1, comprising a particular combination of polymorphisms, differs chemically, structurally, and functionally from a molecule of haplotype 2 and from a molecule comprising a single polymorphism (e.g., PS1). The special technical feature of each haplotype or genotype is the combination of polymorphisms contained therein, which feature is lacking from and not shared with each other haplotype or genotype or with, e.g., a molecule comprising any single polymorphism set forth in the claims. Similarly, with respect to the pairs of polymorphisms of Claim 8, each combination of polymorphisms differs from each other combination and from each of the other combinations discussed above (i.e., haplotypes, genotypes, and single polymorphic sites). Thus, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby, and the claims will be examined only as they read upon the invention of the elected group.

Further, the different methods have different objectives and require different process steps. The haplotyping methods require steps of identifying haplotypes and haplotype pairs to achieve the objectives of haplotyping. The methods of genotyping require steps of identifying a single nucleotide on one gene copy to achieve the objective of genotyping. The methods of predicting a haplotype pair require steps of identifying two polymorphisms in a gene to achieve the objective of "predicting a haplotype pair". The methods of identifying an association requires steps of comparing frequencies of haplotypes in a population to achieve the objective of "identifying an association between a trait" and a haplotype. The methods of assaying for binding activity require steps of assaying for binding activity for candidate agents. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the detection or identification of molecules having the same or common special technical feature, for the reasons discussed above.

It is noted that the polypeptide claim, claim 28, is directed to polypeptides which are fragments of a variant sequence such that the claims do not require the variant is encompassed within the fragment. Thus, this claim reads on a peptide.

The groups comprising polynucleotides, kits, recombinant organisms, polypeptides, antibodies, computer systems and genome anthologies are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids are composed of nucleotides and function in, e.g., methods of nucleic acid hybridization or amplification. These groups are directed to different combinations of nucleic acids which are different from one another and may be employed in different methods. The recombinant organisms are complex organisms that are employed in, e.g. animal research methods. Such organisms cannot be employed as, e.g., probes or primers and they differ in both structure and function from the nucleic acids. The polypeptides differ in both structure and function from either the nucleic acids or the transgenic organisms. The polypeptides are composed of amino acids linked by peptide bonds and arranged in a complex combination of alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. The polypeptides also differ in function, e.g., fusion proteins with an enzymatic functions. The antibodies are composed of amino acids linked by peptide bonds, antibodies are glycosylated and their tertiary structure is unique, where four subunits (2 light chains and 2 heavy chains) associated via disulfide bonds into a Y-shaped symmetric dimer. The

antibodies function in immunoassays. Further the computer systems are composed of, e.g., a CPU, a display device, an input device, etc., as recited in Claim 31, and function in, e.g., methods of electronic sequence comparison. Accordingly, the products differ structurally and functionally from one another. As products of different sets of Groups differ from each other in structure, function, and effect, they do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature".

Continuation of B. FIELDS SEARCHED Item 3:
medline, biosis, caplus, embase, scisearch
CFL1, SEQ ID NO: 65,